Science

28 March 2008 | $10

Gene Regulation

AAAS

Volume 319, Issue 5871

COVER

The many layers of gene regulation in a eukaryotic cell, envisioned as a video game. Transcription in the nucleus (green circle) proceeds to translation in the cytoplasm via genome topology, polymerase pausing, microRNA repression, RNA splicing, and riboswitch regulation. See the special section beginning on page 1781.

Illustration: Carin L. Cain

DEPARTMENTS

EDITORIAL

SPECIAL SECTION

Gene Regulation

INTRODUCTION

NEWS

PERSPECTIVES

>> Editorial p. 1733; for online content, see p. 1727 or go to www.sciencemag.org/cgi/generegulation/

NEWS OF THE WEEK

NEWS FOCUS

CONTENTS continued >>

Science

SCIENCE EXPRESS

www.sciencexpress.org

MATERIALS SCIENCE

Stretchable and Foldable Silicon Integrated Circuits
D.-H. Kim et al.
High-performance, bendable, and stretchable electronic devices are fabricated on an elastic plastic substrate by placing the critical electronic components in the neutral bending plane.
10.1126/science.1154367

APPLIED PHYSICS

Silica-on-Silicon Waveguide Quantum Circuits
A. Politi et al.
Quantum circuits—in which individual photons interfere, entangle, and form logic gates—have been realized on silicon chips.
10.1126/science.1155441

BIOCHEMISTRY

Reconstitution of Pilus Assembly Reveals a Bacterial Outer Membrane Catalyst
M. Nishiyama, T. Ishikawa, H. Rechsteiner, R. Glockshuber
The cell-free formation of the protruberant pilus of a pathogenic bacteria is accelerated by a protein that catalyzes supramolecular assembly without input of cellular energy.
10.1126/science.1154994

GENETICS

Rare Structural Variants Disrupt Multiple Genes in Neurodevelopmental Pathways in Schizophrenia
T. Walsh et al.
Patients with schizophrenia carry multiple small deletions and duplications in their DNA that are associated nonrandomly with neuronal signaling and brain development pathways.
10.1126/science.1155174

LETTERS

BOOKS *ET AL.*

POLICY FORUM

PERSPECTIVES

REVIEW

MATERIALS SCIENCE

BREVIA

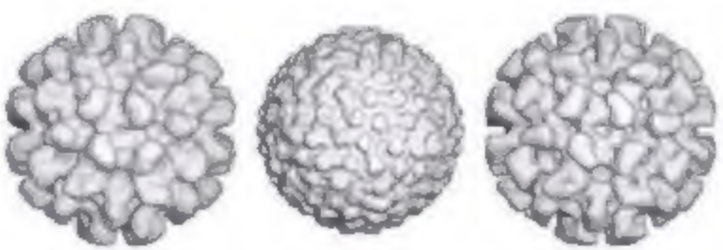
PLANETARY SCIENCE

REPORTS

ASTROPHYSICS

CONTENTS continued >>

REPORTS *CONTINUED...*



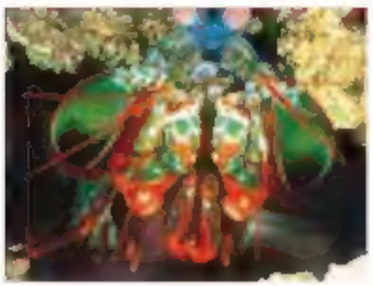
1834

ADVANCING SCIENCE, SERVING SOCIETY

SCIENCE (ISSN 0036-8075) is published weekly on Friday, except the last week in December, by the American Association for the Advancement of Science, 1200 New York Avenue, NW, Washington, DC 20005. Periodicals Mail postage (publication No. 484460) paid at Washington, DC, and additional mailing offices. The title SCIENCE is a registered trademark of the AAAS. Domestic individual membership and subscription ($53 issues: $144 [$74 allocated to subscription]. Domestic institutional subscription ($1 issues): $970. Foreign postage extra: Mexico, Caribbean (surface mail) $55, other countries (air assist delivery) $85. First class, airmail, student, and emeritus rates on request. Canadian rates with GST available upon request. GST #125439122. Publications Mail Agreement Number 1069624. SCIENCE is printed on 30 percent post-consumer recycled paper. **Printed in the U.S.A.**

Change of address: Allow 4 weeks, giving old and new addresses and 8-digit account number. **Postmaster:** Send changes of address to AAAS, P.O. Box 96178, Washington, DC 20090-6178. **Single-copy sales:** $10.00 current issue, $15.00 back issue prepaid includes surface postage; bulk rates on request. **Authorization to photocopy** material for internal or personal use under circumstances not falling within the fair use provisions of the Copyright Act is granted by AAAS to libraries and other users registered with the Copyright Clearance Center (CCC) Transactional Reporting Service, provided that $20.00 per article is paid directly to CCC, 222 Rosewood Drive, Danvers, MA 01923. The identification code for Science is 0036-8075. Science is indexed in the Reader's Guide to Periodical Literature and in several specialized indexes.

CONTENTS continued >>



The eyes have it.

SCIENCENOW

www.sciencenow.org DAILY NEWS COVERAGE

New Form of Vision Discovered
Mantis shrimp eyes can see circular polarized light, which may be used in mating or secret signaling.

Therapeutic Cloning Shows Promise for Parkinson's Disease
Mice treated with their own cells.

What Does a Plant Sound Like?
A computer program reveals how bats find their favorite foliage—and how we can use the same trick.

SPECIAL SECTION

Gene Regulation

SCIENCE SIGNALING

www.stke.org THE SIGNAL TRANSDUCTION KNOWLEDGE ENVIRONMENT

EDITORIAL GUIDE: Focus Issue—Mechanisms of Gene Regulation
J. F. Foley
Multilayered mechanisms control various aspects of gene expression.

PERSPECTIVE: Silent Assassin—Oncogenic Ras Directs Epigenetic Inactivation of Target Genes
X. Cheng
Oncogenic Ras directs a program that epigenetically silences genes that inhibit tumorigenesis.

PERSPECTIVE: NFAT Is Well Placed to Direct Both Enhancer Looping and Domain-Wide Models of Enhancer Function
P. N. Cockerill
Inducible intrachromosomal looping between the tumor necrosis factor–α (TNF-α) gene promoter and two NFAT-dependent enhancers activates TNF-α gene expression.

PERSPECTIVE: SRC-3 Transcription-Coupled Activation, Degradation, and the Ubiquitin Clock—Is There Enough Coactivator to Go Around in Cells?
D. M. Lonard and B. W. O'Malley
The critical factor in estrogen-dependent growth of breast cancer cells appears to be the abundance of the coactivator protein SRC-3.

The hazards of perfectionism.

SCIENCE CAREERS

www.sciencecareers.org/career_development
CAREER RESOURCES FOR SCIENTISTS

Mind Matters: Too Perfect?
I. S. Levine
Perfectionism can diminish productivity, undermine job satisfaction, and damage work relationships.

Mastering Your Ph.D.: Goodbye to All That
P. Gosling and B. Noordam
Once you've said goodbye to the bench, you can take comfort in the opportunities that await.

Educated Woman, Postdoc Edition, Chapter 14: Interview Excursions
M. P. DeWhyse
It's official: Micella is out shopping for a new career.

From the Archives: The Mind Matters Index
I. S. Levine
Learn how to lead a more successful and fulfilling work life.

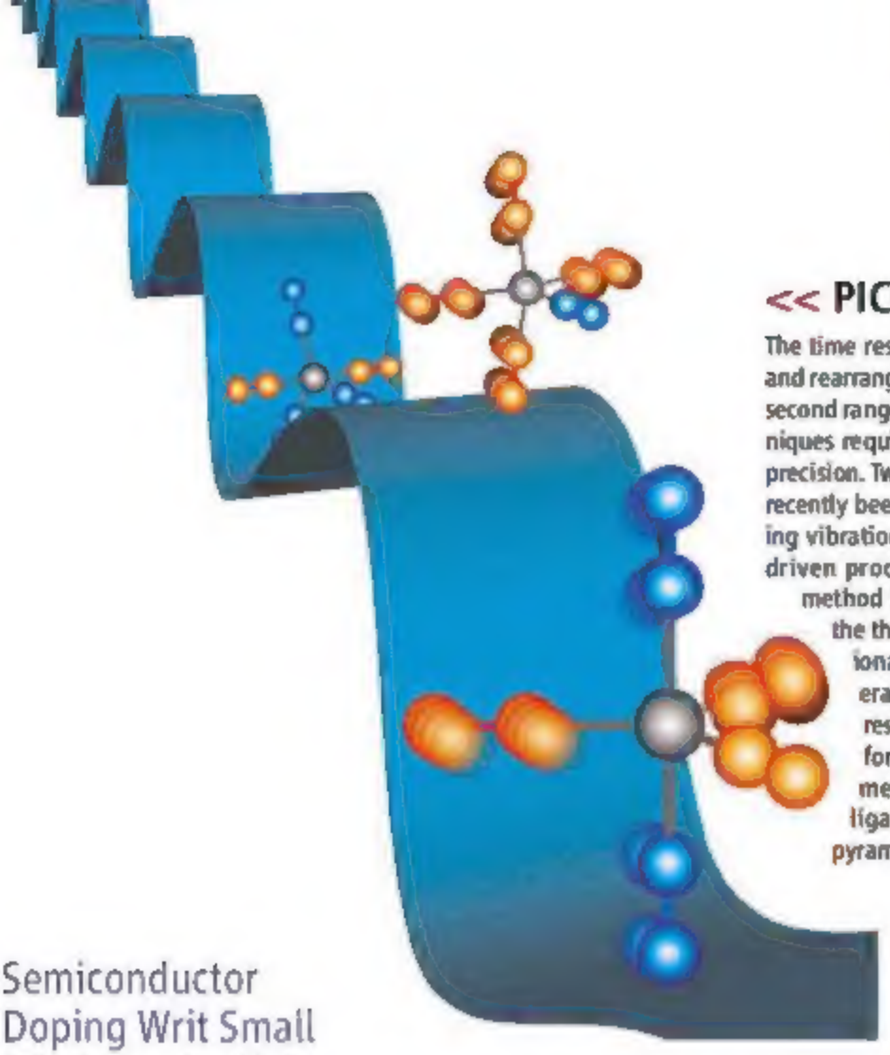
SCIENCE PODCAST

Download the 28 March *Science* Podcast to hear about rule learning by rats, the biomechanical properties of squid beaks, making sense of genome-wide association studies, and more.

www.sciencemag.org/about/podcast.dtl

CREDITS: (SCIENCE NOW) JUSTIN MARSHALL/UNIVERSITY OF QUEENSLAND; (SCIENCE CAREERS) PHOTODISC

Separate individual or institutional subscriptions to these products may be required for full-text access.



<< PICOSECOND PIROUETTE

The time resolution available to track chemical reactions and rearrangements has steadily increased into the femtosecond range of atomic vibrational periods, but most techniques required photoinitiation to achieve the necessary precision. Two-dimensional infrared spectroscopy has very recently been shown to overcome this limitation by tracking vibrational energy migration during rapid thermally driven processes. **Cahoon *et al.*** (p. 1820) apply this method to quantify the transition-state dynamics of the thermally driven ligand rearrangement, or fluxionality, in $Fe(CO)_5$. By modeling the data at several different temperatures and comparing the results with theory, they obtain direct evidence for the long-postulated Berry pseudorotation mechanism in which axial and equatorial CO ligands switch places through a fleeting square pyramidal geometry on a picosecond time scale.

Semiconductor Doping Writ Small

Adding dopants or impurities is a known method to change the electronic properties of semiconductor materials. In theory, it should also be a useful trick for altering the properties of semiconductor nanocrystals or quantum dots. However, as the particle size gets smaller, it becomes increasingly difficult to dope the particles uniformly or to avoid unwanted chemical reactions of dopants that introduce an extra electron or hole into the particle. **Norris *et al.*** (p. 1776) review a number of techniques that have been developed to overcome the challenges of doping nanocrystals.

Careful Hold of a Sharp Tool

The squid beak is an organic hard tissue embedded in soft muscle tissue. **Miserez *et al.*** (p. 1816; see the Perspective by **Messersmith**) question how the squid can use the sharp and rigid beak without causing damage to its own muscle tissue that holds and powers it. They find that there is a gradient in the properties along the beak such that only the cutting end is stiff and hard; the end that is held by the soft tissues is soft and compliant. Gradient materials have been found in nature before, but in this case the authors map and correlate the mechanical properties with the local chemistry. In particular, they find that the stiffness gradient can be tied to mixtures of chitin and a histidine-rich protein family that contains 3, 4-dihydroxyphenyl-L-alanine.

CREDITS (TOP TO BOTTOM): JAMES CAHOON; ROSENBAUND ET AL.

Ionically Driven Membrane Assembly

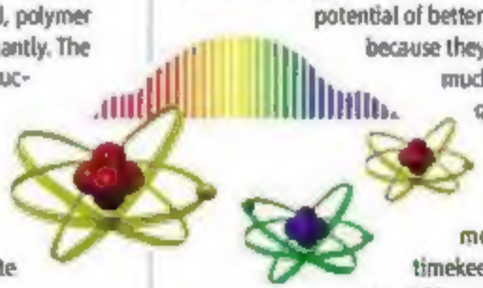
Interfaces can help drive the self-assembly of molecules into larger well-ordered structures. For example, Langmuir-Blodgett films form from amphiphilic molecules at air-water interfaces. **Capito *et al.*** (p. 1812) show that when aqueous solutions of a high molecular weight polysaccharide and an oppositely charged low molecular weight peptide amphiphile are mixed, polymer sacs on the millimeter scale form instantly. The membranes possess a hierarchical structure in which the high molecular weight polymer extends across the membrane. The membranes are self-healing and are robust enough to withstand suturing. Thus, it may be possible to use them to encapsulate cells or other objects.

Evolving Neutron Stars

After a star has spent its nuclear fuel, it may collapse to form a compact neutron star. If this very dense object is rotating quickly, it may then become a pulsar and emit a rotating beam of radio waves that can be detected on Earth as a regular series of radio pulses. Neutron stars can also evolve into other exotic objects called magnetars, which emit x-ray bursts powered by the star's extremely strong magnetic fields. **Gavriil *et al.*** (p. 1802, published online 21 February) have now found a long-sought missing link between these two kinds of neutron stars by examining data from a pulsar in the Aquila constellation. A series of magnetar-like x-ray bursts was observed, and the spectrum of the bursts suggests that the neutron star is changing from a pulsar into a magnetar.

Finely Spaced Clock Ticks

Atomic clocks based on Cs atoms are the standard timekeepers of today, but clocks based on optical transitions of trapped atoms and ions offer the potential of better precision because they operate at much higher frequency (see the Perspective by **Kleppner**). One requirement for any new timekeeping standard is the ability to compare the operation of one remote clock with another. **Ludlow *et al.*** (p. 1805, published online 14 February) make a comparison over an optic fiber line of two different optical clocks (one Sr and the other Ca) 4 kilometers apart, and demonstrate a



Continued on page 1731

Continued from page 1729

fractional uncertainty in the ticking of both clocks of 1×10^{-16} that surpasses even the best Cs-based standards. **Rosenband *et al.*** (p. 1808, published online 6 March) take two different optical clocks based on singly trapped Al and Hg ions and show that the ratio of the frequency of the clocks can be measured with a fractional error of 5.2×10^{-17}. Taking measurements spanning the length of 1 year, they then show that if there are changes in the fundamental constants, measured in terms of the fine structure constant, then such changes are constrained to less than $(-1.6 \pm 2.4) \times 10^{-17}$ per year.

They Are What They Eat

Whether a honeybee larva will become a worker or a queen is largely determined by what it eats—only future queens are fed royal jelly. **Kucharski *et al.*** (p. 1827, published online 13 March) hypothesized that this developmental decision was epigenetically controlled by DNA methylation. Silencing DNA methyl transferase 3 (Dnmt3) mimicked the effects of royal jelly on early developmental processes. When Dnmt3 expression was reduced, even in the absence of royal jelly, larvae developed into queens rather than workers.

Viral Activation Strategy

Flaviviruses, such as dengue, West Nile, or yellow fever viruses, are first assembled as immature, intracellular particles that must be processed and secreted through the secretory pathway. Intracellular cleavage of the precursor membrane protein (prM) in immature particles is essential for infectivity and correlates with virus pathogenicity. **Yu *et al.*** (p. 1834) and **L. Li *et al.*** (p. 1830) provide insight into the structural transitions that occur during virus maturation and how these relate to function. On immature virus, prM undergoes a large conformational change at low pH during transit through the Golgi complex, which exposes its furin cleavage site. The prM protein is cleaved, but remains associated with the virion envelope glycoprotein (E), which masks its membrane fusion loop and prevents premature membrane fusion. On exposure to the neutral pH of the extracellular milieu, prM dissociates from E to give mature virus that can undergo membrane fusion during the infection of a new host cell.

Smell and Repel

Blood-feeding insects are responsible for spreading some of the deadliest infectious diseases. Topically applied insect repellents play a crucial role in protecting humans from these insects. The most widely used of these compounds, DEET, has been around for more than 50 years, but its function is still poorly understood. **Ditzen *et al.*** (p. 1838, published online 13 March; see the 14 March news story by **Leslie**) found in both fruit flies and in the malaria mosquito that DEET acts on the insect olfactory system by inhibiting olfactory neurons that mediate responses to attractive substances. It seems that DEET functions by masking the host odor through blocking odorant receptors that require the olfactory co-receptor OR83b.

The Smell of Fear

Research on classical conditioning has largely concentrated on understanding how an organism learns to associate sensory stimuli with biologically salient events. Can aversive learning directly modify sensory perception of the conditioned stimulus itself? **W. Li *et al.*** (p. 1842) show that classical conditioning can indeed have a direct effect on perceptual discrimination abilities by using odors that exist in enantiomeric forms that cannot normally be distinguished. One form was associated with a negative experience (an electric shock). After repeated exposure, human subjects clearly showed improved discrimination. Conditioning caused representation of the target odors to be reorganized in piriform cortex.

Of Rats and Rules

The learning of rules and the ability to generalize between learned events and novel instances is a fundamental attribute of human cognition. **Murphy *et al.*** (p. 1849) investigated rule learning in nonprimates by asking whether rats can also learn "rules" that are analogous to the rules of grammar—for example, triads like subject-verb-object. Specifically, in a Pavlovian conditioning procedure, rats were trained to discriminate certain triplet sequences from other sequences. The animals' ability to discriminate correct from incorrect sequences developed gradually over a number of training sessions, which suggests that in rats transfer to novel stimuli—learning a rule—is possible.

CREDIT: FEBMILAB

Bruce Alberts is Editor-in-Chief of *Science*.

Shortcuts to Medical Progress?

WE HAVE ALL BEEN TAUGHT THAT THE SHORTEST DISTANCE BETWEEN TWO POINTS IS A straight line. But the same idea has repeatedly proven not to be true for progress in medical research. Why?

More than 80 years ago, in his great book *The Cell in Development and Heredity*, Edmund B. Wilson wrote that "the key to every biological problem must finally be sought in the cell, for every living organism is, or at sometime has been, a cell." The striking modern advances in our understanding of cells, such as those concerning gene expression highlighted in this special section (pp. 1781–1799), make it clear that the chemistry that makes life possible is enormously complex.

I have been part of a team writing a cell biology textbook for 30 years. With each new edition, the authors are repeatedly struck by the fact that scientists still know only a small fraction of what is needed to understand even the simplest bacterial cell. The knowledge gap is of course much greater for scientists trying to come to grips with multicellularity; that is, with the workings of an organism like a fruit fly or a human, in which many billions of individual cells must cooperate to produce an individual. What does this deficit mean for the large-scale ongoing efforts across the globe to use biomedical sciences to improve human health?

The public and the Congress in the United States, including many of the most effective advocates for increased public funding of the biomedical sciences, appear largely unaware of this knowledge gap or of the need to remove it in order to intervene effectively in most human diseases. As in many other nations, there has been a great deal of understandable pressure to increase focus on research that attempts to use what we already know about cells to cure human disease.

I certainly support a head-on assault against diseases wherever feasible. But, as has been repeatedly demonstrated, the shortest path to medical breakthroughs may not come from a direct attack against a specific disease. Critical medical insights frequently arise from attempts to understand fundamental mechanisms in organisms that are much easier to study than humans; in particular, from studies of bacteria, yeasts, insects, plants, and worms. For this reason, an overemphasis on "translational" biomedical research (which focuses on a particular disease) would be counterproductive, even for those who care only about disease prevention and cures.

This issue of *Science* illustrates how progress in basic biological research serves as a major driving force both for advancing knowledge and for spurring new technologies and therapies. Several Perspectives focus on the widespread role of different types of small RNA molecules in controlling the expression of genes. This phenomenon was discovered to be central to human biology only recently, but has already spawned the formation of numerous biotech companies that hope to use this understanding to create new types of drugs to treat humans. It is telling that the entire field developed only because of unexpected results obtained from fundamental research on gene control mechanisms in plants and worms, some of the worm work being recognized with a Nobel Prize in 2006. Without prior fundamental research on these organisms, we would probably have no understanding of the critical roles that small RNAs play in human biology today and therefore no new approach to therapies based on this knowledge.

The private sector energetically pursues translational biomedical research in many areas. But only governments (and a select group of foundations such as the Wellcome Trust and the Howard Hughes Medical Institute) provide the resources needed for pioneering work on fundamental biological mechanisms. With so many mysteries remaining about the incredibly sophisticated chemistry of life, it is certain that future medical breakthroughs will depend to a substantial extent on research on organisms that are much smaller and easier to investigate than ourselves. This counterintuitive fact may need to be much better conveyed to the public if governments are to make sound investments for improving health.

– Bruce Alberts

10.1126/science.1157923

EDITED BY GILBERT CHIN AND JAKE YESTON

BIOMEDICINE

SCOPING OUT COLON CANCER'S VARIED TERRAIN

Colorectal cancer remains a leading cause of cancer deaths in the developed world despite the availability of a well-established screening procedure—colonoscopy—that can detect the disease at an early stage. Although patient noncompliance with screening recommendations is a major contributing factor to the disease's continued prevalence, questions have also been raised about whether conventional colonoscopy can reliably detect the full range of premalignant lesions that eventually progress to cancer. These lesions include not only adenomatous polyps but also morphologically subtle "flat" lesions. Soetikno *et al.* recently reported that flat lesions were present in nearly 10% of 1800 patients screened at a veterans' hospital and that these lesions were five times as likely as polyps to show malignant features.

These findings underscore the importance of ongoing efforts to improve the sensitivity of colonoscopy (the inner surface of a healthy colon is shown above). A progress report on one such strategy is provided by Hsiung *et al.*, who used bacteriophage phage display libraries to identify a small peptide that binds preferentially to human premalignant colonic tissue. When a fluorescein-conjugated version of the peptide was sprayed onto the colon's inner surface and imaged by confocal microendoscopy during routine colonoscopy, it was found to bind to premalignant versus normal cells with 81% sensitivity and 82% selectivity. Although not directly tested in this study, such targeted peptides could in principle be optimized to detect the subtler lesions that are missed by current technologies. — PAK

J. Am. Med. Assoc. **299**, 1027 (2008); *Nat. Med.* **14**, 10.1038/nm1692 (2008).

CREDITS (TOP TO BOTTOM): DR. KESSEL & DR. KARDON/TISSUES & ORGANS; LEE *ET AL.*, *BIOMATERIALS* **29**, 2113 (2008)

CHEMISTRY

Rotaxane Receptor

A rotaxane is a molecular mimic of a wheel and axle—a large ring-shaped molecule that can translate and rotate along a second linear molecule, until its path is blocked bulky end groups. Frey *et al.* examined the application of rotaxanes bearing two macrocyclic groups as receptors in a functional arrangement resembling bookends. Each macrocycle was tethered through a long fused aromatic arm to a Zn porphyrin group, which could be used to bind pyridine molecules. The authors explored the binding of two different guest substrates—one comprising two pyridine rings bonded back-to-back at the 4 position and the other connecting two pyridines through a flexible 10–carbon-atom alkyl bridge—and found that both had nearly the same association constant for insertion into the gap between the Zn centers. Even when the macrocycles were anchored on the axle through copper coordination, the tethered Zn complexes retained sufficient flexibility to bind the guests with similar association constants. — PDS

J. Am. Chem. Soc. **130**, 10.1021/ja7110493 (2008).

BIOTECHNOLOGY

A-Maizing Antibodies

Monoclonal antibodies that block the binding of HIV to cellular receptors have been shown to neutralize the virus in vitro, to protect monkeys from HIV challenge, and to prevent viral transmission through mucosal tissue. But such antibodies can be produced only at high cost and low capacity through expression in mammalian cells and are therefore not optimal for commercial manufacture. Two studies describe the purification of the anti-HIV antibody 2G12 from genetically engineered maize on a large and cost-effective scale. Rademacher *et al.* used a fluorescent marker protein to identify and breed transgenic plants that accumulated a high amount of 2G12 in the seed endosperm, the plant's specialized storage tissue. Ramessar *et al.* purified the antibody from 2G12-expressing maize without using protein A–affinity chromatography, a step typically used for antibody isolation, but toxic if protein A leaches into the final product. Despite differences between mammalian and plant-specific processing of protein-linked carbohydrate, both studies found that glycan modification of maize-produced 2G12 antibodies did not alter antibody binding to the gp120 subunit of the envelope protein of HIV. The HIV-neutralizing properties of mammalian cell– and maize-produced 2G12 were comparable, with the latter being somewhat more potent. Maize-produced 2G12 could be an effective prophylactic mucosal microbicide, and large-scale plant cultivation and prolonged seed storage in the absence of cold temperatures make this method of antibody production economically attractive. — LC

Plant Biotechnol. J. **6**, 189 (2008); *Proc. Natl. Acad. Sci. U.S.A.* **105**, 3727 (2008).

BIOMATERIALS

A Fade-Away Injection

Although a number of drugs can be delivered via a transdermal patch, this pain-free method generally does not match the efficacy of a needle injection. In a hybrid approach, Lee *et al.* fabricated microneedle arrays out of two biocompatible polysaccharides, carboxymethylcellulose (CMC) and amylopectin. A key to making robust needles was pre-concentration of the aqueous polysaccharide solutions

Continued on page 1737

Continued from page 1735

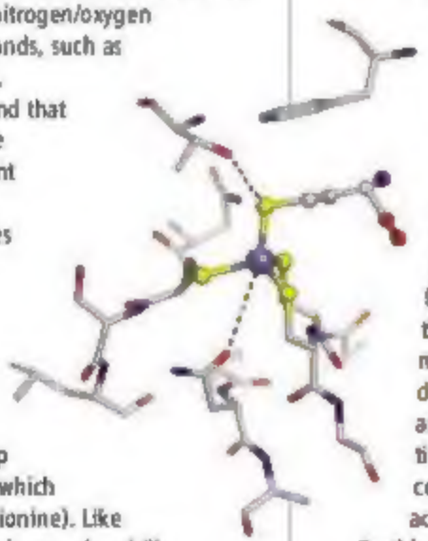
before casting the needles in a polymer mold. The strength of the microneedles depended on shape, because cylindrical CMC needles were not strong enough to penetrate skin, although pyramidal ones were. Drugs could be loaded into the needles themselves, into a backing layer, or both, depending on whether dosing required a bolus shot or a more prolonged delivery. On contact with skin, the microneedles dissolved, releasing the drug and creating pathways for transport from the backing layer. Using lysozyme as a model protein, the authors showed that it was possible to store the arrays for up to 2 months with almost no loss of enzymatic activity. — MSL

Biomaterials **29**, 2113 (2008).

BIOCHEMISTRY

Inversion at Zinc

One of the first things taught in organic chemistry is that a carbon atom can form bonds to four other atoms. Furthermore, these bonds point toward the vertices of a tetrahedron, with the carbon atom located at its center. The consequent variety of carbon-based molecules is, of course, the basis for biochemistry and, in particular, the biological families of macromolecules—lipids, nucleic acids, carbohydrates, and proteins. Zinc can contribute either to the catalytic power or the structural integrity of proteins and usually binds in tetrahedral fashion to the sulfur or nitrogen/oxygen atoms of its four ligands, such as cysteine or histidine.



Koutmos *et al.* find that the zinc atoms in the cobalamin-dependent and -independent methionine synthases MetH and MetE undergo an inversion in their geometry as these enzymes mediate the transfer of a methyl group onto homocysteine (which reaction yields methionine). Like a tetrahedral carbon in a nucleophilic substitution reaction, the zinc atom releases one of its ligands (glutamate or aspara-gine, respectively) as it reaches to make contact with the sulfur in homocysteine; this motion resembles an umbrella turning inside out on a windy day. The unanticipated flexibility of an active-site metal fits nicely with recent thinking about the importance of intrinsic protein motions for enzyme catalysis. — GJC

Proc. Natl. Acad. Sci. U.S.A. **105**, 3286 (2008).

CREDIT: KOUTMOS ET AL., PROC. NATL. ACAD. SCI. U.S.A. 105, 3286 (2008)

CELL BIOLOGY

Activity at the Pore

Throughout the nucleus, chromosomal regions localize to particular subcompartments, such as the nuclear periphery, that correspond with gene activation or repression. For example, genes positioned near the lamin-rich regions of the nuclear envelope in *Drosophila* generally correlate with transcription repression. Relatively less is known about spatial regulation in mammalian cells. Brown *et al.* have used biochemical mapping and visualization techniques in human cells to elucidate genomic organization at the nuclear pore complex. Upon treatment with a histone deacetylase inhibitor, extensive nuclear reorganization resulted, with an increased association of the nuclear pore protein nucleoporin 93 with genomic features and transcription factors characteristic of transcriptionally active genes. Hence, the nuclear pore plays an important part in the regulation of genes in the human genome. — BAP

Genes Dev. **22**, 627 (2008).

GEOLOGY

Copper in the Crust

The concentration of most metals is extremely low in Earth's crust. For copper, the most important ores globally are the porphyry deposits, which form when copper and other metals are concentrated in fluids emanating from and flowing through shallow bodies of magma in the crust. Rapid cooling of these fluids, or a sudden change in their chemistry as they interact with rocks or other fluids, leads to precipitation of copious amounts of metals and metal sulfides. Their formation has been episodic throughout Earth's history, and many ore deposits, as they form at depths, have been eroded. Kesler and Wilkinson account for the formation, uplift, and erosion of porphyry copper deposits to provide a global accounting of these deposits through Earth's history and assess the remaining available resource. They estimate that about 0.25% of the copper in Earth's crust has actually been concentrated in ores, and that about two-thirds of the more than 100,000 ore deposits that have formed over Earth's history have been eroded and recycled. At current consumption rates, they estimate that there is about a 5000-year supply of copper remaining in the Earth. — BH

Geology **36**, 255 (2008).

www.sciencemag.org

Science

1200 New York Avenue, NW
Washington, DC 20005
Editorial: 202-326-6550, FAX 202-289-7562
News: 202-326-6581, FAX 202-371-9227
Bateman House, 82-88 Hills Road
Cambridge, UK CB2 1LQ
+44 (0) 1223 326500, FAX +44 (0) 1223 326501

Subscription Services For change of address, missing issues, new orders and renewals, and payment questions: 866-434-AAAS (2227) or 202-326-6417, FAX 202-842-1065. Mailing addresses: AAAS, PO Box 96178, Washington, DC 20090-6178 or AAAS Member Services, 1200 New York Avenue, NW, Washington, DC 20005

Institutional Site Licenses please call 202-326-6755 for any questions or information

Reprints: Author inquiries 800-635-7181
Commercial inquiries 803-359-4578

Permissions 202-326-7074, FAX 202-682-0816

Member Benefits AAAS/Barnes&Noble.com bookstore www.aaas.org/bn; AAAS Online Store http://www.aproute.com/aaas/ (code MKB6); AAAS Travels Betchart Expeditions 800-252-4910; Apple Store www.apple.com/epp/store/aaas; Bank of America MasterCard 1-800-833-8727 (priority code FAAJYU); Cold Spring Harbor Laboratory Press Publications www.cshlpress.com/cshl/site/aaas.htm; GEICO Auto Insurance www.geico.com/aaas; Hertz 800-654-2200 (CDP # 437); Office Depot http://www.officedepot.com/portalLogin.do; Seabury & Smith Life Insurance 800-424-9883; Subaru VIP Program 202-326-6417; VIP Moving Services http://www.vipmayflower.com/domestic/index.html; Other Benefits, AAAS Member Services 202-326-6417 or www.aaasmember.org

science_editors@aaas.org (for general editorial queries)
science_letters@aaas.org (for queries about letters)
science_reviews@aaas.org (for returning manuscript reviews)
science_bookrevs@aaas.org (for book review queries)

Published by the American Association for the Advancement of Science (AAAS). *Science* serves its readers as a forum for the presentation and discussion of important issues related to the advancement of science, including the presentation of minority or conflicting points of view, rather than by publishing only material on which a consensus has been reached. Accordingly, all articles published in *Science*—including editorials, news and comment, and book reviews—are signed and reflect the individual views of the authors and not official points of view adopted by AAAS or the institutions with which the authors are affiliated.

AAAS was founded in 1848 and incorporated in 1874. Its mission is to advance science and innovation throughout the world for the benefit of all people. The goals of the association are to foster communication among scientists, engineers and the public; enhance international cooperation in science and its applications; promote the responsible conduct and use of science and technology; foster education in science and technology for everyone; enhance the science and technology workforce and infrastructure; increase public understanding and appreciation of science and technology; and strengthen support for the science and technology enterprise.

Information for Authors

See pages 634 and 635 of the 1 February 2008 issue or access www.sciencemag.org/about/authors

Looking Within

Corpus, a Dutch museum that opened its doors last week on the road between Amsterdam and The Hague, offers visitors a 50-minute trip through a giant human body, starting in the knee. Organs such as the brain and the womb serve as minitheaters in which three-dimensional videos and sound effects—such as a sucking sound in the lungs—help explain everything from pregnancy and wound healing to sneezing.

The creature sports both male and female reproductive organs but is definitely "human," says its creator and director, Henri Remmers, a self-described "concept developer" who also helped set up an archaeological theme park in the Netherlands.

After visits around head and guts, visitors exit from the brain and into an information center that promotes healthy lifestyles. Supported by corporations, medical groups, and the government, the €20 million building—which from the outside resembles a seated giant—doubles as a convention center.

Chimp Corridor Proposed

An Iowa conservation group hopes to salvage a tiny group of chimpanzees in Rwanda by creating a forest "corridor" to connect them with a larger population.

The Gishwati Forest used to cover 100,000 hectares, but logging and farming have shrunk it by 90%. A colony of up to 20 chimps has been hanging on in isolation for decades. The Great Ape Trust in Des Moines is now collaborating with Rwanda's forest authority to plant trees near the isolated chimps, collecting seeds from chimp feces to make sure they are trees that the animals like to use. They plan to cover 80 hectares this year and hope to eventually create a continuous stretch of forest that will allow the population to merge with some 800 chimps that live in two national parks about 50 kilometers away.

Such corridors are relatively untried. But the trust's conservation director, Benjamin Beck, cites a previous successful 10-year project in Brazil involving golden lion tamarins. Beck estimates the cost in Rwanda at a few million dollars, mostly to compensate landowners. Locals are skeptical right now, he acknowledges, but he hopes that continued meetings will help win them over.

Science Off the Air

Nearly half of Americans cannot name a "role model" scientist, living or dead. And only 11% can come up with the name of a living one, according to a survey released last week by the Museum of Science and Industry in Chicago, Illinois. And whom do they think of most often? Bill Gates and Al Gore. Each was named by 6% of the sample, on a par with Albert Einstein. Most respondents also reported that citizens' ignorance of science is "a detriment to our nation."

A possible source of the problem emerged from another study released last week by the Pew Research Center for the People and the Press. It found that for every 5 hours of U.S. cable television news, only 2 minutes are devoted to science or the environment. By contrast, the same period contains 10 minutes of celebrity news and nearly half an hour on crime.

Given the priorities in their major news outlets, "it's not surprising that in polls, few Americans rank climate change or the environment as a top political priority or even a major national problem," says Matthew Nisbet, a social scientist at American University in Washington, D.C.

MOTETS, MONKS, AND MORTAR

An architectural historian has taken a choir to Venice to determine how much Renaissance architects and composers shaped each other's work. Last spring, with acousticians and musicologists, Deborah Howard of Cambridge University in the U.K. led an experimental public concert tour on which the Choir of St. John's College, Cambridge, performed Renaissance works in 11 Venetian churches and monasteries, including the San Marco basilica.

Recordings, as well as audience reactions, indicated that complex polyphonic pieces reverberated too much throughout large spaces such as the basilica but sounded right in San Marco's smaller ducal chapel. Monastery chapels were the best settings for resonant but straightforward chants. And humbler parish churches adorned with sound-damping tapestries were suited to simple hymn singing. "Each church did generate the kind of acoustic that was appropriate" to its needs, says Howard, showing that architects designed with acoustics in mind.

Composers also probably tailored their work to specific buildings, says Howard, who presented her findings at this month's Cambridge Science Festival. For example, the team found compositions calling for a double choir that in a reverberating space such as San Marco would achieve a "surround sound" effect. "We suppose that many musicians compose their work having in mind a very particular kind of place," says applied physicist Francesco Martellotta of the Polytechnic University of Bari, "but in this case, it is clearly documented."

CREDITS (TOP TO BOTTOM): CORPUS; ERYOSA OTA/REUTERS/CORBIS; LAURA MORSTH

EDITED BY YUDHIJIT BHATTACHARJEE

POLITICS

RAISING HACKLES. During her many years surveying brown bears in Montana and surrounding states, Katherine Kendall has stood her ground in the face of growling grizzlies. The 56-year-old U.S. Geological Survey (USGS) biologist has demonstrated the same steeliness against Arizona Senator John McCain, the presumptive Republican presidential nominee, who has repeatedly cited Kendall's $5 million project in a campaign to eliminate what he regards as wasteful earmarking by Congress.

Kendall's large-scale DNA analysis of bear hair, which was first added to the USGS budget in 2003, provides a nonintrusive way to do a census of the bear population in the Northern Continental Divide Ecosystem, with an eye toward determining the best way to manage the species. But McCain has ridiculed it as useful only for investigating bear paternity suits and dumpster crimes.

"It's not a comfortable feeling to have that kind of spotlighting in an unfavorable light," Kendall says about the salvos McCain launched during recent speeches in South Carolina and in a November 2007 television campaign commercial. "We were very careful about making every penny count." Kendall hopes to publish the results sometime this year.

MOVERS

GENE SEEKER. A Nigeria-born genetic epidemiologist who has led studies on why some ethnic groups suffer higher rates of certain diseases has been chosen to head a new $1.7 million Center for Genomics and Health Disparities at the U.S. National Human Genome Research Institute in Bethesda, Maryland. Charles Rotimi, 50, had been directing a human genome center at Howard University in Washington, D.C., one of the country's leading minority-serving institutions.

Rotimi is working on the first genome-wide association study of an African-American population, looking for DNA involved in diabetes, hypertension, and obesity. He plans to continue studies of diabetes in West Africa and China and join with Duke University scientists to study genetic differences in responses to drugs. He says moving to the NIH campus will allow him to expand collaborations and do high-risk research without the constant pressure to obtain outside grants. Rotimi brought along five workers from his lab at Howard and expects to build a total staff of 10 to 12.

Marine biologist **John Burris** has been named president of the Burroughs Wellcome Fund, an $800 million foundation in Research Triangle Park, North Carolina, that supports biomedical research. Burris, who directed the Marine Biological Laboratory in Woods Hole, Massachusetts, during the '90s, has served as president of Beloit College in Wisconsin since 2000. He starts his new job in July, succeeding Enriqueta Bond.

ON CAMPUS

FELLED, NOT. Brian Karpes was shot five times by the gunman who killed five students and wounded 18 others during last month's shooting at Northern Illinois University (NIU) in DeKalb. But the geology graduate student survived—and so did his plans to present a poster at this month's Lunar and Planetary Science Conference outside of Houston, Texas.

Karpes, 27, was sitting in on a 150-student oceanography class as a teaching assistant when the shooter began firing a shotgun and three handguns from the auditorium stage. Karpes was hit in the head, twice in his left side, and once each in his right hand and left biceps. He spent a week in the hospital. Despite a broken writing hand, Karpes was able to assemble a poster based on his cataloging of impact craters around the equatorial region of Saturn's plume-spouting moon Enceladus. And after doctors said he couldn't travel, an NIU professor agreed to carry the poster to Houston. It was the only one among hundreds at the meeting to be signed by dozens of well-wishers.

Got a tip for this page? E-mail people@aaas.org

CREDITS (TOP TO BOTTOM): DEREK BERTIN; SOURCE: BRIAN KARPES; MAGGIE BARTLETT/NIU

Three Q's >>

Helmut Schwarz, 64, a molecular chemist at the Technical University of Berlin, took over as president of the Alexander von Humboldt Foundation in Bonn in January. The foundation awards hundreds of fellowships each year to researchers from around the world to work in Germany. This year it will begin a program to lure foreign scientists to German universities with 10 5-year, €5 million professorship awards.

Q: What can the foundation do to be more effective?

It is clear that our competition has increased, and other organizations pay much better than we do. It will be important to significantly raise the amount we award fellows [for example, $40,000 per year for postdoctoral fellows] so that we can make sure that no one chooses to go somewhere else based on money.

Q: What do you hope the Humboldt Professorships will achieve?

The award is something new. It's for a person plus an institution. A person must be nominated by the potential host university. The Max Planck Society recruits 30% of its top researchers from abroad, and we want to make it possible for a university to recruit top talent, too. But the universities have to make sure that when the 5 years are through, it's possible for the person to stay.

Q: You have said that the professorships should go to researchers who are playmakers. What do you mean by that?

In recent years, we have become a bit provincial at German universities. We want the Humboldt professors to set something in motion that has an infectious effect on the whole university.

Cooperating rooks
1746

Q&A with China's health minister
1748

GLOBAL CHANGE

Roads, Ports, Rails Aren't Ready for Changing Climate, Says Report

A federal study released this month documents the significant impact that climate change is expected to have on the U.S. transportation system. Its conclusion, says Henry Schwartz, the former head of one of the country's largest highway engineering firms, is "a pretty damning tale of what could happen."

The 3-year effort, led by the U.S. Department of Transportation and including outside experts as well as climate scientists, focused on a roughly 80-km-wide strip along the Gulf Coast region from Mobile, Alabama, to Houston, Texas, that is home to 10 million people (climatescience.gov/Library/sap/sap4-7/final-report/). It found an expected sea-level rise of 122 centimeters over the next 50 to 100 years—an estimate reflecting the midrange of previous global forecasts that was tweaked to factor in as much as 81 cm in subsidence in some areas—would permanently flood nearly a third of the region's major roads. Some 72% of ports in the region would be at risk, the report concludes, and a majority of roads and 29 airports would likely experience major flooding during major storms. In addition, the report says that a temperature rise of between 0.5° and 2.5°C would cause railroad lines to buckle, require more sturdy driving surfaces, and boost the cost of road maintenance by increasing the strain on repair crews.

Scientists say the study is the most rigorous effort thus far to quantify how climate change could impact vulnerable U.S. infrastructure at the local level. "Transportation professionals ... by and large haven't looked very seriously at global warming," says Schwartz, a former president of engineering giant Sverdrup Civil (now Jacobs Engineering Group in Pasadena, California). Schwartz is also chair of a panel of the U.S. National Academies' National Research Council (NRC) that has tackled the same issue on a national scale. Its report, released last week, calls for similar partnerships between climate scientists and transportation planners in other regions (nap.edu/catalog/12179.html).

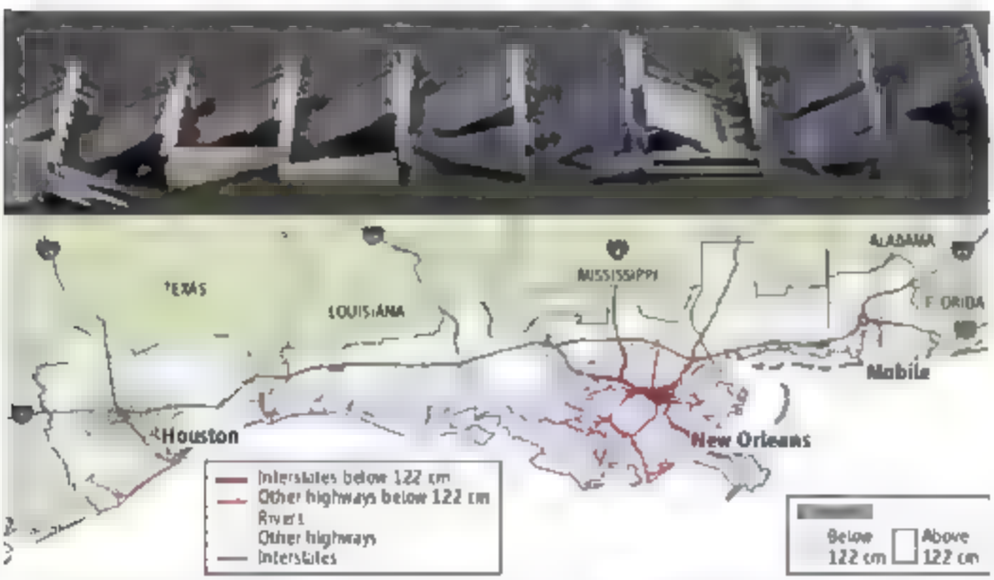
The Gulf Coast study examined how the region's climate might change over the expected life of major transportation facilities, a period that extends for the next half-century and beyond. The panel extracted data from 21 global climate models that project a temperature increase for the region of from 1° to 2°C by 2050. Researchers also took into account geologic data, global forecasts on sea-level rise, and ground subsidence of as much as 0.8 cm per year. Transportation planners then overlaid the forecasts onto thousands of kilometers of roads, ports, railways, and airports.

The results point to a looming disaster. "Based on these levels, an untenable portion of the region's road, rail, and port network is at risk of permanent flooding," the federal report says. Hurricanes and the rising ocean are already destroying barrier islands, which blunt the coastal impact of incoming storms. The Gulf Coast study also calls for more drainage and alternative traffic-handling schemes to cope with increasingly severe rain storms, as well as better evacuation routes for hurricanes expected to grow more intense.

The two reports acknowledge their limitations. The federal study admits that its precipitation projections are open to question and that more rapid polar melting could wreak greater havoc on transportation systems. To address such gaps in the research, the NRC study called for more regional climate studies, short-term prediction tools to foresee disruptions such as storm-flooded roads, and more help from the federal government to local planners trying to batten down the hatches.

Both reports recommend new approaches to infrastructure investments. In the past, says Schwartz, engineers decided how high to make a highway overpass or how deep to make a drainage ditch by using "some preset design standards" based on historic variations in rainfall, temperature, and other climate variables. The changing climate patterns now require planners to devise a standard for each project based on the probability of more severe weather events. Additional federal studies are expected to focus on how climate change would affect a particular Gulf Coast facility, be it a port, highway, or coastal city.

Alan Clark, a transportation planner for the greater Houston, Texas, area who participated in the study, plans to take shifting climate patterns into consideration in ▶



Road hazards. New study (*above*) shows what a 122-cm rise in sea level would do to the Gulf Coast's major roads, already vulnerable to major storms such as Hurricane Katrina in 2005 (*top*).

CREDITS (TOP TO BOTTOM): M. SPENCER GREEN/AP PHOTO; SOURCE: CAMBRIDGE SYSTEMATICS ANALYSIS OF U.S. DOT DATA

preparing a development plan for the Houston-Galveston area. In particular, says Clark, climate change should be factored into decisions on whether to build a proposed second bridge connecting the rapidly growing Galveston Island to the mainland. Aside from the usual factors such as cost, environmental impact, and the population to be served, Clark notes, officials also need to think about whether continued coastal development is prudent in the face of rising sea levels and worsening storms. "The real question is whether we should be going down this route at all," he says.

Transportation managers in several southern and western states say they would like to undertake similar studies of their regions but that money is an obstacle. John Zamurs of the New York State Department of Transportation, who advised on the Gulf Coast effort and has tried to launch an equivalent effort for his state, says, "We're still struggling to find the funding."

Putting the issue off would be a mistake, says Schwartz. The Missouri-based Sverdrup Civil he once headed has built billions of dollars' worth of water, highway, rail, and bridge projects across the country. But he says that the challenges to infrastructure will be more complex and the costs greater as both the seas and the mercury rise. —ELI KINTISCH

CLIMATE CHANGE

Study Fingers Soot as a Major Player in Global Warming

Climate-change authorities long ago tagged carbon dioxide public enemy number one. Now, there may be a new number two: tiny particles of black carbon, or soot. According to a new analysis reported online this week in *Nature Geoscience*, climate scientists are concluding that reports such as last November's assessment by the Intergovernmental Panel on Climate Change (IPCC) may seriously underestimate black carbon's role in global warming. The good news is that—unlike reductions in greenhouse gas emissions—reducing the release of large amounts of black carbon worldwide would have immediate effects.

Although the error bars on the new measurement are large, "the effects of black carbon are definitely stronger than what the IPCC estimates," says Mark Jacobson, an atmospheric scientist at Stanford University in Palo Alto, California, who was not involved in the study.

The IPCC report noted that black carbon is a strong absorber of sunlight but downplayed its impact because the haze it produces occurs regionally rather than globally. IPCC estimated that, at current levels, black carbon warms the atmosphere by 0.2 to 0.4 watts per square meter ($W m^{-2}$), considerably below the value of 1.66 $W m^{-2}$ for CO_2. But in their new analysis of a wide variety of recent data, Veerabhadran Ramanathan of the Scripps Institution of Oceanography in San Diego, California, and Gregory Carmichael of the University of Iowa in Iowa City suggest that black carbon warms the atmosphere by as much as 0.9 $W m^{-2}$, enough to vault it over the impact of other climate-warming gases such as methane, halocarbons, and tropospheric ozone.

CREDIT: REZA/WEBISTAN/CORBIS

Black carbon comes from sources as diverse as the burning of grasslands in Africa and the rainforest in Brazil, diesel emissions from trucks in North America, cooking fires burning coal in China, and cow dung in India. The soot wafts high into the atmosphere, often in thick brown clouds that block some sunlight and absorb significant solar radiation.

Ramanathan says previous conservative estimates of black carbon's warming effect overlooked key factors—most importantly, the interaction between black carbon and other particles in the atmosphere. "Black carbon doesn't exist by itself," says Ramanathan, "It's always mixed with other aerosols," such as sulfate particles, and other organic combustion byproducts. Many of those other aerosols reflect light, increasing the chances that it will be absorbed by nearby flecks of soot. Black carbon high in the atmosphere also absorbs light reflected by Earth's surface and clouds. Because most climate models don't adequately represent such effects, they often underestimate how much reflected light soot absorbs, Ramanathan says.

If studies by Ramanathan and others turn out to be right, that's both good and bad news for policymakers. On the downside, the plethora of sources of black carbon will make it hard to cut emissions. On the other hand, the particles circulate in the atmosphere for only about a week before falling back to Earth. So concerted efforts to reduce biomass burning in the Amazon, cut diesel emissions, or convert cooking stoves to biogas or even solar power could have a powerful impact far more quickly than changes in CO_2 emissions.

Public health researchers have long sought to curb black carbon because inhalation of smoke from wood fires is thought to contribute to as many as 400,000 deaths a year in India alone. Now, Jacobson says, climate change research "underscores the need to control black carbon even more." —ROBERT F. SERVICE

BEHAVIORAL ECOLOGY

Smart Birds Lend a Beak for Food

Corvids—including ravens, crows, and rooks—are among the smarty-pants of the bird world. Now scientists report that rooks, like chimpanzees, can cooperate in food-getting tasks.

Researchers have long sought to demonstrate cooperation in animals to probe the cognitive underpinnings of this behavior. These insights may help explain how cooperation evolved so spectacularly in humans.

In 2006, researchers showed that chimps could work together to bring a food-laden platform within reach of their cage (*Science*, 3 March 2006, p. 1297). Now, a team of researchers in the United Kingdom has shown that rooks are capable of the same behavior, even though they are vastly removed evolutionarily from chimps. "All the things that have been popping up in primates seem to be popping up as well in social birds," says Harvard University primatologist Marc Hauser.

Scientists have already demonstrated that corvids will share food and help a pal in a fight, much as chimps do. To see how well rooks would cooperate in a food-getting task, Amanda Seed, Nicola Clayton, and Nathan Emery at the University of Cambridge in the U.K. rigged up a platform containing dishes of worms and cooked egg yolk. A string passing through loops on the platform provided beak holds that the birds could use to drag the food into their cage. Because the two ends were too far apart for one bird to pick up, two individuals—one at each end—had to pull simultaneously. If one pulled by itself, the string would slip through the loops.

Strings attached. Rooks will work together if it's convenient.

The researchers exposed eight pairings of eight rooks to the challenge. In 60 trials per pair, all were able to pull in the eggs and worms. But as with chimps, the pairs who were most "tolerant"—that is, got along with each other the best as evidenced by behaviors such as feeding from the same dish—picked up the trick the fastest. The most tolerant pair reeled in the food platform in 63% of the tries, whereas the least tolerant had a score of 20%, the team reports online 26 March in the *Proceedings of the Royal Society*.

But unlike chimps, the rooks did not have the insight to realize they would be better off waiting for help. The researchers monitored individual birds placed in a room with the test platform. The bird could either pull fruitlessly on one end of the string or wait a minute or more for a partner to push through a one-way flap and help with the job. All the birds pulled on a string without waiting for aid.

Nor did the birds show a preference when given a choice between pulling the string when they could reach both ends and pulling it when the ends were too far apart. This result confirmed that the birds were not able to discriminate between situations that required help and those that did not.

It's a "wonderful paper," says Brian Hare of Duke University in Durham, North Carolina, who did the chimp studies. He says it supports the idea that tolerance is crucial for the evolution of cooperation. But the rooks' cognitive skills may not be quite as advanced as those of the chimps. The "delay" task, he notes, shows that rooks "don't really have a solid understanding of needing the other individual" to get the goodies.

–CONSTANCE HOLDEN

CLINICAL RESEARCH

NIH Reports Breach of Patient Records

The theft last month of a laptop with patient data from a clinical trial under way at the U.S. National Institutes of Health (NIH) in Bethesda, Maryland, is fueling broader concerns about privacy. The computer, stolen from the locked car trunk of an NIH scientist, has led to a tightening of NIH policies on safeguarding patient information and raised concerns about practices at U.S. universities.

The incident, first reported this week by *The Washington Post*, involved Andrew Arai, head of the cardiovascular magnetic resonance imaging (MRI) section at the National Heart, Lung, and Blood Institute (NHLBI). The laptop contained names, birth dates, and some clinical data for 2500 patients taking part in a 7-year study of cardiac MRI. The files did not include Social Security numbers, phone numbers, or addresses. The laptop was password-protected but not encrypted. Last May, the government began requiring laptops containing sensitive data to be encrypted following the 2006 theft of a laptop containing personal information on millions of Veterans Affairs (VA) patients.

Nearly 1 month after Arai reported the off-campus theft to NIH officials, an internal NHLBI committee approved a letter to patients notifying them of the incident. A privacy advocacy group expressed concern about the delay, noting that there is no federal requirement that patients be informed of a security breach. "It exposes a weakness of our national standards," says Deven McGraw of the Center for Democracy and Technology in Washington, D.C.

NHLBI Deputy Director Susan Shurin says that "this particular breach was unlikely to cause harm." Still, she says, NIH now realizes it needs to tighten its policies. NHLBI has already taken a first step by forbidding intramural researchers from storing identified patient data on a laptop. It also will clarify that patient information should be stored on network servers, Shurin says, and that only coded data may be downloaded onto desktops.

After the VA laptop was stolen, NIH cautioned its grantee institutions to use encryption and other methods to protect sensitive data. But it's not clear whether that advice is being followed. "It's incredibly difficult to enforce on an investigator-by-investigator basis," Shurin says. David Korn of the Association of American Medical Colleges in Washington, D.C., says "it wouldn't surprise me" to learn that some researchers are carrying unencrypted patient data home on a laptop, adding, "I just think our community has to step up to another level of security concerns."

–JOCELYN KAISER

CREDITS: CHRISTOPHER BIRD (INSET) AMANDA SEED

INFECTIOUS DISEASE

Elusive Pathogen Cornered at Last

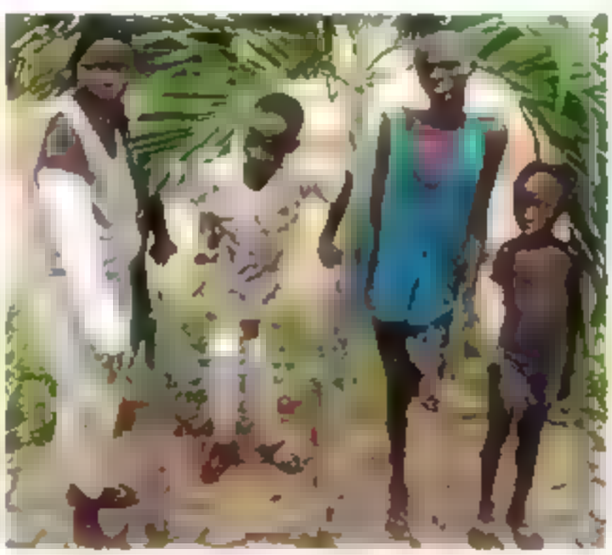
Françoise Portaels has finally answered the question she began to study in 1969 as a Ph.D. student in the Congo. She and her colleagues have discovered an environmental hiding place for the bacteria that cause a devastating disease known as Buruli ulcer. As they report in the March 2008 issue of *PLoS Neglected Tropical Diseases*, Portaels, now a microbiologist at the Institute of Tropical Medicine in Antwerp, Belgium, and her team have cultured the disease-causing organism, *Mycobacterium ulcerans*, from an aquatic insect from Benin. The accomplishment comes at the 10th anniversary of the World Health Organization's (WHO's) Global Buruli Ulcer Initiative and marks a major milestone in efforts to understand and control one of the world's most neglected tropical diseases.

Australian doctors first described Buruli ulcer and *M. ulcerans* in 1948. Although it belongs to a well-known genus that includes the bacteria responsible for tuberculosis and leprosy, the life cycle of *M. ulcerans* has remained mysterious. The infection begins as an innocent-looking bump under the skin that, in several months, opens into a painless lesion. The ulcer gradually spreads over the body, demolishing skin, soft tissue, and sometimes bone, leaving behind severe scars that cause both immobility and social stigma. The disease affects several thousand people each year, mostly in Africa and Australia. The infection is not transmitted person to person but always occurs near slow-moving water, leading researchers to suspect that aquatic insects might harbor and transmit the disease. But without a live culture as evidence no one could be sure.

Ten years ago, Portaels resumed the search she had abandoned after her frustrating dissertation. "The big problem is that *M. ulcerans* is a very slow grower," she explains. It takes 6 weeks to cultivate a colony of *M. ulcerans* even from a human lesion in which it is abundant and active.

To keep fast-growing competitors from overrunning *M. ulcerans* in culture, Portaels's team used an unconventional growth medium: mice. They ground up five different water bugs collected from Benin and Togo and treated the samples to kill everything except mycobacteria. Then they injected the cultured mycobacteria into the paws of mice, where harmless species, not suited for life in a mammalian host, slowly perished but *M. ulcerans* took hold. After passing through three different mice over the course of about 2 years, an *M. ulcerans* lineage from a single water strider began to grow in culture. Nearly 40 years after her first trip to the Congo, Portaels had, for the first time, isolated *M. ulcerans* from the environment.

"It's a significant contribution," says Kingsley Asiedu of WHO's Buruli Ulcer Initiative. But there's a catch. The water strider, *Limnogonus hypoleucus*, doesn't bite people, so it can't transmit the disease. Pamela Small, a microbiologist at the University of Tennessee, Knoxville, and an author of the study, suspects that *M. ulcerans* thrives in another aquatic niche and had to pass through several links in the food chain before infecting the water strider. "Eventually, we will find where this organism is replicating," she says.



Scarred. Bacteria that cause the disfiguring Buruli ulcer disease have been cultured from an African water strider.

CREDIT: WHO, NATIONAL BURULI ULCER CONTROL PROGRAMME BENIN

Portaels, Small, and dozens of other researchers will assemble in Geneva, Switzerland, on 31 March at WHO's 10th annual meeting on Buruli ulcer. Asiedu says the initiative's most notable success has been the development of antibiotic remedies for the disease—but much remains to be done: There is a need for earlier diagnosis, faster treatment, and better understanding of transmission. For the latter, Asiedu says Portaels's work is a boon. "This is the first time it's been proven beyond any doubt that *M. ulcerans* [occurs] in aquatic insects," he says. "It will stimulate interesting discussions."

–ELSA YOUNGSTEADT

Last Collider Standing

Next month, U.S. particle physicists will be down to their last particle smasher. On 7 April, the Stanford Linear Accelerator Center in Menlo Park, California, will shut down the PEP-II collider, 5 weeks after Cornell University's CESR collider took its last data. That leaves only the Tevatron at Fermi National Accelerator Laboratory (Fermilab) in Batavia, Illinois, still running—and it will shut down in 2010 at the latest.

The United States is not pulling out of particle physics; more than 1200 U.S. physicists are working on experiments that will run at the Large Hadron Collider (LHC), which will power up this summer at the European laboratory for particle physics, CERN, near Geneva, Switzerland. That's more than any other country in the world, but Aboulhassan Jawahery of the University of Maryland, College Park, worries that after LHC, "it's a question" of how the U.S. program can thrive without a domestic collider. American physicists hope the answer lies in hosting the proposed International Linear Collider at Fermilab. But there's competition from Europe and Japan, and Congress has reduced funding this year for research on that multibillion-dollar machine.

–ADRIAN CHO

Fund Urged for Tropical Diseases

Developed countries should set up a special fund to fight neglected tropical diseases (NTDs) such as intestinal worms, schistosomiasis, elephantiasis, and river blindness, five influential researchers from around the world say in the current issue of *PLoS Neglected Tropical Diseases*. While the world is waging a multibillion-dollar attack against three major infectious killers through initiatives such as the Global Fund to Fight AIDS, Tuberculosis, and Malaria, NTDs have received scant attention—a "tragic oversight," the group writes. The researchers, who hope government leaders will launch the fund at this summer's G8 summit in Toyako, Japan, put the cost at $2 billion for the first 5 years.

–MARTIN ENSERINK

Gates Is Rainmaker for Drought Research

The Bill and Melinda Gates Foundation announced last week a joint $47 million donation with the Howard G. Buffett Foundation to help develop drought-tolerant corn in Africa. The International Maize and Wheat Improvement Center in Nairobi will work with technology from Monsanto and BASF. The African Agricultural Technology Foundation plans to distribute seeds royalty-free.

–[illegible]

CHEN ZHU INTERVIEW

China's Modern Medical Minister

BEIJING—China's minister of health, Chen Zhu, has an enormous responsibility: to improve medical care for 1.3 billion people in a rapidly developing country. Now he is busier than ever. A reshuffle this month put Chen's ministry in charge of the State Food and Drug Administration, whose previous chief was found guilty last year of taking bribes—a crime for which he was executed.

If Chen is feeling pressure to clean up China's scandal-ridden food and drug industries, he doesn't show it. Perhaps that's because the hematologist, who earned a Ph D in Paris, is not your typical bureaucrat. For starters, he's one of only two Chinese ministers (the other being science minister Wang Gan) who are not members of the Communist Party. And Chen is keeping a hand in research, with a paper about a traditional Chinese medicine (TCM) preparation for acute promyelocytic leukemia that appeared online earlier this month in the *Proceedings of the National Academy of Sciences*.

Investigating this form of leukemia is how Chen made his scientific name in the 1990s, when he led a team that unraveled the molecular mechanism of arsenic-based TCM therapy for the disease. His honors include membership in both the Chinese Academy of Sciences and the U.S. National Academy of Sciences.

"There is a need for a specialized funding channel for medical research. This idea has gotten the support of the top-level policymakers." —CHEN ZHU

Chen took a break from the annual sessions of the National People's Congress and the National Committee of the Chinese People's Political Consultative Conference to discuss the tension between Western medicine and TCM, a plan to strengthen medical research, and the more open research atmosphere in China since the SARS outbreak in 2003.

–RICHARD STONE

Q: You have many challenges providing adequate health care in a country so large and diverse and with a growing gap between rich and poor. How can science help?

Z.C.: Even if you have a health care system that covers everyone, if you don't have the right strategy, the system will be difficult to sustain because of an aging population and chronic diseases. What's happening in the developed world today is China's tomorrow.

We need to focus on disease prevention. To be effective, we need more epidemiological studies. On one hand, we need to strengthen the community medical health care system, and on the other hand, we need to strengthen the science of prevention and early intervention.

China's health care policy has three pillars: One is prevention, the second is taking the countryside as the emphasis, and the third is an equal position for Western medicine and traditional Chinese medicine.

Q: How can the belief system of TCM be reconciled with the Western approach to medicine?

Z.C.: First of all, you need to have proven clinical efficacy. And you need good models at the organism level, at the cellular level, and at the molecular level. To establish clinical efficacy, I encourage my colleagues in

ACADEMIC RESEARCH

Saudi Start-Up Hopes Grants Will Buy Time

Universities often throw money at top scientists they want to hire. But an unusual graduate university rising in the Saudi Arabian desert (*Science*, 8 June 2007, p. 1409) has come up with a new twist on that formula: Let the scientists remain where they are but ask them to share their expertise and contacts to help the school launch its own research programs.

Backed by a $10 billion endowment, officials at the King Abdullah University of Science and Technology (KAUST) this month awarded 12 researchers up to $10 million each over 5 years for research at their home institutions (kaust.edu.sa). The scientists are free to use the money as they see fit, with no immediate returns to KAUST. The one string is that the awardees will spend a minimum of 3 weeks a year talking shop at KAUST, which will open in September 2009 at a site bordering the Red Sea.

The announcement comes on the heels of deals struck with three U.S. universities and a British institution, worth $25 million over 5 years, to help KAUST develop its curricula, recruit faculty, and establish its research agenda. Those agreements involve individual departments at Stanford University, the University of Texas, and the University of California, Berkeley, as well as Imperial College London. This month, KAUST also announced it had hired Fawwaz Ulaby, an electrical engineer and former administrator at the University of Michigan, Ann Arbor, as provost.

Half the awards in the Global Research Partnership (GRP) initiative announced last week will go to U.S. scientists, with the rest sprinkled around the globe. The 11 men and one woman range in age from their early 30s to their 60s and work in one or more of four broad areas—energy and the environment, materials science and engineering, biosciences, and computational sciences—important to Saudi Arabia's development. "It will have a huge impact on my research," says synthetic organic chemist Brian Stoltz, 37, who was recently appointed to a chaired professorship at the California Institute of Technology in Pasadena. The $10 million award, he says, "is probably three to five times what I get now for my lab." Stoltz says he's intrigued by KAUST's attempt to "start from scratch and build a high-quality university."

The winners were drawn from a pool of 60 schools invited by KAUST to apply, divided equally among North and South America, Europe and Asia, and the rest of the world. A blue-ribbon panel led by Frank Press, former president of the U.S. National Academy of Sciences, ranked the proposals and forwarded its recommendations to KAUST, which made the final selection.

Some of the grantees hope to spend more than the required 3 weeks a year at KAUST.

CREDIT: ING HAN GUAN/AP PHOTO

TCM to organize multicenter studies. I tell them, this is not Western practice; this is universal practice!

I oppose the idea that TCM is something sacred, something you cannot dissect.

Q: Some TCM practitioners argue that you just have to trust that it works

Z.C.: If it works, there must be some material basis, there must be mechanisms.

Q: Is there truth to the rumor that you hope to create a Chinese version of the U.S. National Institutes of Health (NIH)?

Z.C.: This is one of my dreams. Maybe something similar but not identical to NIH. The world has different models. NIH is a very successful one.

We do think there is a need for a specialized funding channel for medical research. This idea has gotten the support of the top-level policymakers. The Ministry of Science and Technology agrees, and the Ministry of Finance agrees. We are now preparing an interdepartmental council.

A consortium of research institutions will form an intramural program. We have the Chinese Academy of Medical Sciences, the Chinese Academy of Traditional Chinese Medicine, and some institutes of the Chinese Academy of Sciences. And we have the Chinese Academy of Military Sciences, which is of excellent performance in emerging infectious diseases. They played a major role in SARS. We also need to provide more support to the extramural teams, basically the universities, hospitals, medical centers. Further, we need to support risk-taking projects.

Q: Will there be additional funding?

Z.C.: Hopefully! We have a mutual understanding with the Ministry of Science that we will not take a piece of their cake. But hopefully the Finance Ministry can support us by adding some more resources.

Q: How will the interdepartmental council work?

Z.C.: We have some preliminary ideas. To encourage fair play, we cannot say that all the teams at these institutions should be supported. We need to select centers of excellence. The extramural program will be mainly based on a call for proposals to university research teams and medical centers. But we must have guidelines. My idea is a combination of bottom-up and top-down approaches. For a country like China, we need sometimes the top-down approach, particularly for issues like food safety and emerging infectious diseases.

Q: Have concerns about food safety been resolved to your satisfaction?

Z.C.: Not yet. The overall level has been increased, I have to say. We have evidence that food-borne disease has declined over the past several months. Now we need a long-term mechanism. And we need a quicker response. Capacity building for detection and information gathering and dissemination, an early warning system—all these things are very important. The future council will be dealing with these issues, and we need to mobilize investigators. So sometimes we need the top-down approach.

Q: What lessons have been learned since the SARS outbreak?

Z.C.: In the health care system, we need a concept and structure for crisis management. This is the lesson we learned. Of course, at large, we need to balance economic growth with social development. Our leaders' outlook of scientific development was initiated just after SARS. So in one sense, SARS incubated this concept.

Q: During the SARS outbreak, government officials pressured researchers not to report data. Is the atmosphere better now?

Z.C.: Transparency is really the key issue. I get a daily report on infectious diseases and public health emergencies nationwide through an Internet-based reporting system. We get reports directly from regional centers for disease control. They don't need to get approval from local authorities.

We now have the biggest Internet-based disease-reporting system in the world. We are extending it to other areas, to cover chronic illnesses. We need transparent policies and open reporting of major diseases.

whereas others say a heavy workload will make it difficult to do more than the minimum. Although details of the interactions have yet to be worked out, several investigators mentioned the possibility of two-way exchanges between KAUST faculty and students and members of their labs in addition to guest lectures, workshops, and symposia.

Despite the generous funding and minimal requirements, only 38 of the 60 institutions took up KAUST's offer, and only two-thirds of those submitted the maximum of two proposals per school. Physical chemist Bengt Nordén, 63, of Chalmers University of Technology in Göteborg, Sweden, says university officials debated whether to participate at all before submitting his winning proposal for $10 million. "The laws and culture of Saudi Arabia are somewhat controversial, and it was not obvious that we should say yes," he explains. "But after reading about their commitment to Western standards of academic freedom, including treating men and women equally, we decided that even if we don't agree with every government policy, it is better to communicate than to not communicate."

CREDITS (LEFT TO RIGHT): COURTESY OF B. NORDÉN; A. TRAMONTANO

Mixing oil and money. Sweden's Bengt Nordén and Italy's Anna Tramontano are part of the first class of KAUST grantees.

Anna Tramontano, a 50-year-old computational biologist at the University of Rome, "La Sapienza," who will receive $5 million, says she had "no concerns about participating, because science shouldn't consider anything else" besides the quality of the research to be carried out. Although she sees her selection as "a signal" that the Saudi government respects the contributions of women, what really excites her is the opportunity to apply vastly increased computing power to comparative analyses of the multiple human genomes that will soon be available. "I never thought my proposal would be chosen, but it was fun to think about what I could do. And now I have the chance."

In the coming weeks, KAUST will announce the winners of two other competitions, for large university-based research centers and for postdoctoral fellowships. Officials expect to make at least two additional rounds of GRP grants, although the pool of eligible universities may vary.

–JEFFREY MERVIS

Science by the Masses

By offering prizes on behalf of clients seeking scientific and engineering help, an Internet company called InnoCentive has gathered a virtual workforce of 135,000 problem solvers around the world.

Until 15 months ago, Harvey Arbesman knew relatively little about amyotrophic lateral sclerosis (ALS), the fatal neuromuscular disease commonly known as Lou Gehrig's disease. That would quickly change for the practicing dermatologist and professor of clinical epidemiology at the University at Buffalo in New York. While reading about how the online encyclopedia Wikipedia and others benefit from a concept called crowd-sourcing, Arbesman was drawn to a company named InnoCentive and a biomedical contest it had posted on the Internet.

Crowd-sourcing typically involves allowing a mass of people to help a company or group accomplish its goals. Wikipedia, for example, allows people to write, correct, and update its encyclopedia entries. InnoCentive exploits that strategy in the scientific realm, posting technical and theoretical challenges online and offering awards ranging from $5000 to $1 million for a solution. What caught Arbesman's eye was a set of challenges run on behalf of an ALS patient group called Prize4Life, including a $1 million prize for discovering a validated biomarker that tracks the progression of the disease, and smaller awards for promising ideas.

Practicing what he preaches—Arbesman and his wife have a company devoted to creative problem-solving in medicine—Arbesman began reading everything he could find on ALS. He soon came across the little-studied observation that ALS patients, when immobilized during the end stage of their disease, rarely develop bedsores, even though similarly paralyzed people, such as stroke victims, almost always do. "Is there a clue there?" wondered Arbesman.

The hunt for ALS biomarkers has concentrated on molecules in blood, urine, and cerebrospinal fluid—so far without success. But the dermatologist ultimately submitted a proposal for monitoring skin changes in ALS patients. Impressed, Prize4Life awarded him a $15,000 prize for a ALS biomarker concept. "It's a fascinating idea," says Hiroshi Mitsumoto, director of the ALS center at Columbia University, who is now working with Arbesman to test the skin of about 40 to 50 people with the condition.

Such stories provide the public relations pitch for InnoCentive's radical approach to scientific problem solving. With a success rate greater than 35% for the 600-plus challenges the company has posted since 2001, on topics such as synthesizing a tuberculosis drug and building bricks cheaply in developing countries, InnoCentive CEO and President Dwayne Spradlin boasts that his company now has more than 135,000 "solvers" worldwide addressing his clients' intractable problems. "The prize-based model can be better, faster, cheaper" than traditional in-house research efforts, he says.

InnoCentive has drawn a diverse crowd of scientists and engineers into its virtual workforce. About 40% of those who register to see challenge summaries have Ph.D.s. Karim Lakhani of Harvard Business School in Boston, who was given access to InnoCentive's data on challenges from 2001 to 2004 and also surveyed about 350 of its solvers, has found that curiosity and pride motivate them as much as the prize money. He suggests that the company's crowd-sourcing approach reflects a "broader trend of democratization of science." As the United States and Europe churn out Ph.D.s, and countries such as China and India dramatically expand their scientific

CREDIT: IMAGE100.COM/CORBIS

capabilities, more and more people with science training exist outside the traditionally elite research universities. "Many people have the skills and talents to solve science problems," says Lakhani.

Success from failure

The drug company probably doesn't like to think of it this way, but Eli Lilly's research and development failures gave birth to InnoCentive. Frustrated that a massive scientific staff couldn't solve certain problems in drug synthesis or development, Eli Lilly officials took a chance and spun off InnoCentive in 2001. It began posting challenges from Eli Lilly and other drug companies, though InnoCentive allows the companies seeking help to remain anonymous if they wish. Spradlin notes that some companies were initially hesitant to reveal problems or tip off competitors, but solvers interested in the full technical demands must sign a confidentiality agreement.

Although some contests ask for an actual product—an intermediate chemical in a drug-synthesis effort or a particular mutant yeast strain, for example—many others are more theoretical. Take a drug company seeking new ideas for how to treat obesity. The InnoCentive client determines if an award should be given for any submitted solution, and InnoCentive then helps it obtain the intellectual-property (IP) rights to the idea from the solver in return for the prize money. InnoCentive brags that it has negotiated IP transfer for more than 99% of solutions recognized by its clients.

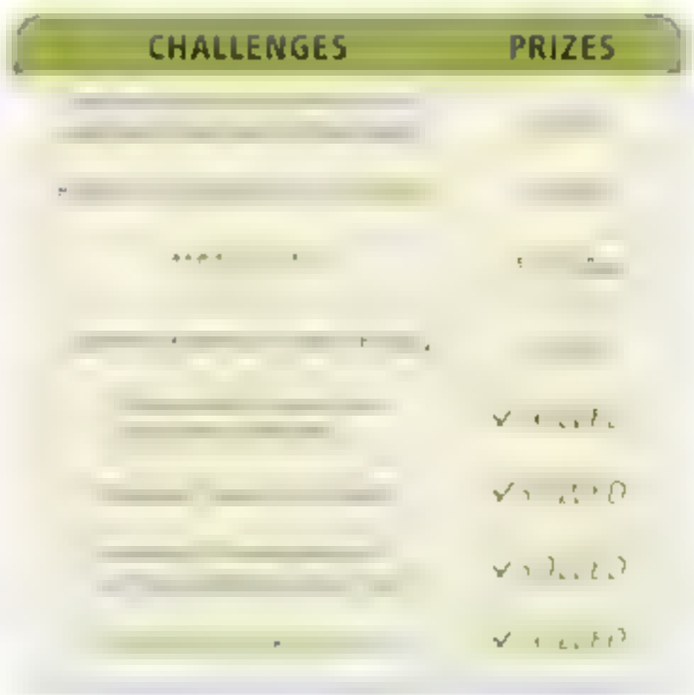
InnoCentive, originally owned and subsidized by Eli Lilly, has now raised more than $9 million in venture capital as an independent company. InnoCentive, which basically maintains a Web site, makes money by charging clients annual fees of up to $100,000 to post challenges and manage the exchange that follows. InnoCentive also earns a percentage of the prize money, sometimes equal to the award itself, if a company agrees that a solver has found an answer. Spradlin says the company has so far awarded about $2.6 million for 200-plus solutions, most coming within 2 to 4 months of a challenge being posted.

InnoCentive soon caught the attention of the business world, with several publications dubbing it the eBay of innovation. Major clients from the pharmaceutical, materials, and chemistry industries have come onboard. The publicity has fueled rapid growth: 50 to 100 new solvers register every day, says Spradlin. They hail from 175 countries and, aside from traditional science Ph.D.s, include technicians, students, and engineers.

More than 50% of registered solvers now come from Russia, India, and China. InnoCentive has even signed agreements with the Chinese and Russian national science academies; instead of preventing their scientists from answering challenges, these organizations now promote InnoCentive. As motivation for Russian universities, for example, a solver's academic department can get 10% of any award, says Spradlin.

Doing good

Recently, InnoCentive has begun to branch out from its pharmaceutical heritage, offering nonprofit organizations and others a platform.



SOURCE: INNOCENTIVE

Take Prize4Life, which a wealthy ALS patient founded in 2006 to accelerate ALS drug discovery. Given the small number of people affected, drug companies ignore the disease, in part because the lack of a good biomarker makes evaluating treatments a challenge, says Nicole Szlezak of Prize4Life. In the hope that prizes could serve as a "lighthouse" that illuminates new research and ideas, the ALS group put up its $1 million offer and smaller awards through InnoCentive. It now plans to use the company for more prizes, including a $2.5 million prize for an ALS treatment that works in mice (*Science*, 8 February, p. 713).

The Oil Spill Recovery Institute (OSRI), established by Congress after the 1989 *Exxon Valdez* oil tanker disaster, also signed up with InnoCentive, seeking new ideas for cleaning up oil spills in Arctic and sub-Arctic waters. The institute has posted several challenges and recently awarded $20,000 for an idea for promoting the flow of oil; cleanup crews store oil from spills in floating barges but often have trouble emptying the near-frozen liquid. "We had a selection of very good ideas to choose from. We ended up with one we were very satisfied with," notes Scott Pegau of OSRI in a promotional video InnoCentive has posted in an online appeal for new solvers on the video site YouTube. "If it was easily solved by the oil industry, it would have been solved."

The solver in this case was John Davis, a U.S. consultant who holds a master's degree in chemistry. Although he had no background in the oil industry, Davis had experience pouring concrete. He suggested inserting into the barges pneumatic vibrators that the concrete industry uses to keep its material flowing. "I love trying to solve problems," gushes Davis on his own YouTube video. The chemist hopes to visit OSRI in Alaska and says he will use some of his prize money to finance environmental cleanup and soil remediation research.

Responses to other challenges have been mixed, says Pegau, and even Davis's award-winning idea has yet to be turned into an actual solution. Some solvers have no concept of the difficulty of working in Alaska, Pegau says. But he's largely sold on InnoCentive, even if others remain cautious about its unusual approach. "We are learning the types of questions that may make a good challenge," he says. "Other organizations have been waiting to see how our experience is going."

One big name hasn't been so hesitant. In December 2006, the Rockefeller Foundation in New York City agreed to a partnership with InnoCentive in which it pays the company to run challenges for various groups seeking innovations for poor and vulnerable people. Tom Kruer, a Kentucky-based mechanical engineer, has won two such challenges. A $15,000 prize, offered by Rural Innovations Network (RIN) in Chennai, India, called for a way to quickly change a spice mixer from grinding dry grains to processing chili, coriander, and other spices with high moisture content. Attracted by the philanthropic nature of the challenge, Kruer in about 25 hours designed an interchangeable blade apparatus that should be low-cost and easy to use in developing countries.

As a consultant, Kruer estimates he might have charged a client $50,000 for the same

project. RIN "got a solution for $15,000 and employed 260 solvers. I think that's groundbreaking," he says. Kruer spent much more time designing a new device for making clay bricks, that earned him $20,000 from a group called GlobalGiving in Washington, D.C.

Motivations

InnoCentive solvers frequently comment that their solutions were trivial or obvious, which suggests that having the right background is key to addressing a client's "unsolvable" problems. Laurie Parker, a postdoc in an organic synthesis lab at the University of Chicago in Illinois, says it recently took her less than a day to write a submission for an InnoCentive client seeking new ways to make libraries of polypeptides. The challenge was a "perfect fit" for her background—and her proposal won a $5000 prize.

Spradlin says that InnoCentive essentially expands the knowledge base of its clients, often resulting in solutions already devised in other fields or countries being applied to a new problem. He recounts that a pharmaceutical company was stuck scaling up production of a key molecule. Within 3 weeks of posting the challenge, a Russian protein crystallographer had pointed the company to a solution already in the public domain—and thus free of any IP restrictions. "There's real breakthrough science that comes from having people around the world working on a problem," says Spradlin.

Christian Hedberg, a chemical biology postdoc at the Max Planck Institute of Molecular Physiology in Dortmund, Germany, had a similarly easy time making some money. The pharma giant Johnson & Johnson was developing a new tuberculosis drug but needed a better way of producing a supply that was purely one of the two stereoisomers. "When I saw the challenge, I just smiled, because I directly knew how to solve it. It took me three evenings to write it up. I'm just used to fight[ing] with delicate chemical problems on a daily basis," says Hedberg, who also does some pharmaceutical consulting for companies.

Hedberg says his past consideration of similar issues for other molecules meant that the "the solution was already there. … It's nice to see that my extensive reading of the chemical literature over 10 years finally pays off in terms of real applications and not only academic publications."

Hedberg cannot discuss his solution, as he transferred IP rights to the company. Still, he admits being surprised to see such a challenge posted by InnoCentive. "I think it's strange that a major pharma company cannot solve this kind of problem," he says.

"I love trying to solve problems."

—JOHN DAVIS, A CHEMIST WHO APPLIED HIS EXPERIENCE WITH CONCRETE TO CRUDE OIL.

Shake it up. A $20,000 award went to an idea for using pneumatic vibrators to help get near-frozen oil out of cleanup barges.

Prized research

Although prize incentives for research and technological needs are enjoying a resurgence (*Science*, 8 February, p. 713), there's been little research on the topic. In that regard, InnoCentive, which has kept detailed records about every challenge, is a gold mine. A self-confessed "data-miner," Spradlin has noted many patterns, such as that Chinese and other Asian solvers register to look at a lot of different challenges but are much more reluctant to submit answers than, say, Russian colleagues.

For his research, Lakhani joined his Harvard colleague Lars Bo Jeppesen and two InnoCentive scientists to sift through the company's data on 166 challenges listed by 26 companies. They also conducted an online survey of solvers, both those who had winning proposals and those who did not. All in all, about 80,000 scientists from 150 countries reviewed those challenges, and 49 were solved, a rate that Lakhani considers impressive given that most of the problems stumped well-funded research and development companies.

Trying to understand why certain problems got solved, Lakhani found that the more diverse the pool of solvers, the greater the odds of a solution. And after surveying winning solvers, his group concluded that the further a challenge was from a person's field of interest, the more likely they were to solve it. Lakhani initially found this puzzling but now says that "you really need to have a different perspective" on a challenge.

The online survey also proposed more than a dozen motivations for tackling an InnoCentive challenge and asked solvers to rate the importance of each. Although prize money was among the top incentives, it wasn't number one: that was the enjoyment of solving a puzzle. "I'm always looking for a challenge to solve. It's part of my nature," Kruer says. Enhancing one's skills was also a motivation. Parker has examined more than 200 challenges but only submitted an answer twice. "At the very least, I learn something each time," she says.

Teaming up

InnoCentive has ambitious plans to add new features to its business, such as auctioning solutions to industry or allowing solvers to network and work together. Spradlin notes with amusement that one solver in India has already assembled his own team of scientists and engineers and merely manages the submissions of answers. The company discovered this outsourcing, which is perfectly legitimate, because the Indian solver was so much more prolific than anyone else.

Still, ideas are cheap. InnoCentive's clients and solvers admit that transforming them into actual solutions isn't necessarily easy. Since Prize4Life awarded him the $15,000, Arbesman has largely been left on his own to test the skin biomarker concept—and he has a November 2008 deadline to meet if he wants to earn the $1 million prize for a validated biomarker. A dermatologist colleague at Columbia University put Arbesman in contact with the ALS center there, which led to the pilot project, but Mitsumoto notes that larger studies will require much more money.

Although InnoCentive cannot yet claim a blockbuster discovery such as an ALS biomarker or treatment, many feel that the company's approach does have a role to play in tapping the global talent pool of scientists and engineers. "This is not the universal solver for all science problems, but it should be in the portfolio of approaches," says Lakhani.

–JOHN TRAVIS

CREDITS (TOP TO BOTTOM): COURTESY OF JOHN DAVIS; COURTESY OF ALASKA CLEAN SEAS

ENERGY

Weighing the Climate Risks Of an Untapped Fossil Fuel

As the energy industry hungrily eyes methane hydrates, scientists ponder the fuel's impact on climate

VIENNA, AUSTRIA—A recent workshop* on methane hydrates felt like a powwow of 19th century California gold prospectors, looking ahead to both riches and peril. Sizing up the prize, Arthur Johnson, a veteran geologist of the oil industry who is now an energy consultant based in Kenner, Louisiana, predicted that "within a decade or two, hydrates will grow to 10% to 15% of natural gas production," becoming a more than $200 billion industry. And the peril? "The worst-case scenario is that global warming triggers a decade-long release of hundreds of gigatons of methane, the equivalent of 10 times the current amount of greenhouse gas in the atmosphere," said David Archer, a climate modeler at the University of Chicago in Illinois. Although no current model predicts such an event, said Archer, "we'd be talking about mass extinction."

When methane molecules become locked in atomic cages of water called clathrates, they form icy chunks that ignite when lit. These solids form wherever methane encounters water at high pressure and low temperature. The necessary conditions reign in permafrost and in some sea-floor sediments, forming a "ring around the bathtub" on continental slopes. This exotic fuel was discovered by the Soviet petroleum industry more than 3 decades ago, but even a few years ago many doubted its commercial potential (*Science*, 13 February 2004, p. 946). After several successful pilot drilling studies and heavy research investment over the past 4 years, says Johnson, "the question now is not whether industry will exploit hydrates but how soon."

Considering the skyrocketing price of oil, the answer seems to be soon, says one of the workshop organizers, Nebojša Nakićenović, an energy economist here at the International Institute for Applied Systems Analysis (IIASA) outside Vienna. "And yet hydrates are absent from most of the climate discussions," he says, "and virtually absent from the IPCC fourth assessment report," last year's 1000-page tome by the Intergovernmental Panel on Climate Change (*Science*, 11 May 2007, p. 812). The goal of the IIASA workshop was to bring together researchers from all the different fields that touch hydrates—from chemistry and economics to climate impact—to get an "interdisciplinary perspective" on the uncertainties.

"It's clear that one of our biggest knowledge gaps is figuring out the distribution," says Michael Riedel, a marine geophysicist at McGill University in Montreal, Canada. "We still don't know how much there is in the world, not even within an order of magnitude."

Another crucial gap is the flux of methane, which drives hydrate formation over time. The largest amounts of methane hydrates are thought to reside in sub–sea-floor sediments. In a newly built sea-floor–monitoring network called NEPTUNE off the western coast of Canada, Riedel is part of a team studying methane-spewing vents to get a handle on their flow rate and marine chemistry. Where the conditions are just right, methane hydrates form caps over pockets of such gas. These not only are sweet spots for those who want to tap hydrates for energy but also represent a major worry for climate modelers.

"If the sea floor warms up by a few degrees Celsius, the most vulnerable hydrates will melt, and then you're going to get a massive release of methane," says Euan Nisbet, a marine geologist at Royal Holloway, University of London. That warming and release is expected to take centuries or even millennia even in the most extreme climate scenarios. Riedel says the methane bubbles from sea-floor vents are sponged up by the ocean water. But if a methane release were large and shallow enough, it would reach the atmosphere, says Archer. What is unclear is whether the climate system has methane-driven positive feedback mechanisms that could lead to abrupt climate change.

Johnson threw cold water on the scenario of a massive release of submarine hydrate-trapped methane to the atmosphere. Most hydrate deposits found so far "are as deep as a kilometer below the sea floor," he says, "and they aren't going anywhere." Walter Oechel, an ecologist and carbon-cycle expert at San Diego State University in California, doesn't find the "doom-and-gloom scenarios" very likely either. "The real story for me is hydrates as yet another chronic contributor to greenhouse gas emissions," he says.

Great balls of fire! When methane meets water under high pressure and low temperature, it forms icy chunks that burn when lit.

Others considered methane hydrates part of a greenhouse gas solution. A plan proposed by Vladimir Yakushev, a geologist at Gazprom, the world's largest natural gas corporation, based in Moscow, involves simultaneously extracting methane and methane hydrates while pumping liquefied carbon dioxide into the underground spaces left behind. Researchers also discussed the idea of using hydrates for electricity generation or even manufacturing on the spot. "We have to try to make it carbon-neutral if we're serious about climate change," says Nisbet.

The overarching question of whether methane hydrates should play a major role in climate change debate was up for grabs. Considering the workshop discussions, "the methane hydrate issue is one risk that shouldn't drive policy considerations at the moment," concludes Brian O'Neill, an IPCC author and climate modeler at the National Center for Atmospheric Research in Boulder, Colorado. "There are bigger fish to fry." But Neil Hamilton, director of the International Arctic Programme for the World Wildlife Fund, based in Oslo, Norway, says, "It's absolutely shocking that hydrates have gotten so little attention." The risk of a massive methane release, however unlikely, "is reason enough for very serious concern," he says. More meetings like these are clearly needed. **–JOHN BOHANNON**

CREDIT: KAZUHIRO NOGI/AFP/GETTY IMAGES

*"Vulnerability and Opportunity of Methane Hydrates Workshop," IIASA, 13–14 March 2008.

GENETIC RISK

With New Disease Genes, A Bounty of Questions

New techniques, including genome-wide associations, are identifying new disease risk factors; researchers are uncertain what they mean—and what to advise patients

Scientists trying to understand the factors that lead to breast cancer have a wealth of new leads to follow up. During the past couple of years, they have identified an alphabet soup of variations in DNA sequences—CHEK2, FGFR2, TNRC9—that appear to increase a woman's risk of getting breast cancer. But these new finds come with a host of questions—in particular, whether to test women for these genetic variations and what to tell those who carry them. It's been "a matter of uncertainty about how exactly to advise people" on this, says Mark Robson, a breast oncologist and clinic director of the clinical genetics service at Memorial Sloan-Kettering Cancer Center in New York City. For the most part, Robson says, he doesn't.

Counseling women about mutations in the widely cited *BRCA1* and *BRCA2* genes makes sense, says Robson, because they can increase breast cancer risk by as much as nine times—a risk "so high that it clearly exceeds most people's threshold for action." But what to do with a gene variant that shifts breast cancer risk from 13% to 16% over a woman's lifetime? Or one that puts diabetes risk at 9% instead of 7%? By comparison, environmental effects can have a much bigger impact. For example, heavy smokers in their 50s have a 6% risk of dying of lung cancer over 10 years compared with about 0.2% for nonsmokers.

In the last year or so, questions about what to do with the flood of data have taken on new urgency. Genome-wide association (GWA) scans—which survey the genomes of people with a particular disease and compare them with the genomes of those without—are turning up dozens of DNA variations that boost risk only modestly. The results have generated enormous excitement among researchers, long frustrated by their inability to find variants driving common adult diseases. But with the data come more questions.

Researchers are finding that even with all the new details from GWA studies, much of the canvas remains obscure. For example, the function of most GWA variants hasn't been determined. It's also not known whether different variants that increase risk slightly for a disease might interact with or add to one another to increase risk substantially. Nor is it clear how the variants might contribute to disease mechanisms or treatment. Will a public that's apparently hungry for genetic knowledge incorporate low-risk data into their lives? These and related questions—many of which came up at a meeting earlier this month at the U.S. National Institutes of Health in Bethesda, Maryland—have leaders in the field wondering how best to apply recent findings, where to focus the next round of studies, and how to convey often sketchy data to the public.

Risk tolerance

Some concerns about the public's reaction have already eased. Two years ago, behavioral epidemiologist Colleen McBride and human geneticist Lawrence Brody, both at the National Human Genome Research Institute, began examining how healthy individuals respond to disease risk information. They offered volunteers in Detroit, Michigan, the chance to learn whether they carried deleterious variants for eight health conditions, including diabetes, colon cancer, and osteoporosis. Because the variants are common, virtually everyone was expected to harbor at least a couple. Those monitoring the study's safety "were really worried, literally, that people were going to jump off bridges" when they learned that their risk of disease was increased, says McBride.

Among the 300 or so who have participated, that hasn't happened—quite the opposite. "They're not having big emotional responses," says McBride. The researchers are tracking the volunteers to see whether the information affects decisions to reduce disease risk, such as seeking out a smoking-cessation program or consulting with a nutritionist.

Behavioral specialists have shifted from worrying about the devastating effects of learning about these new genetic risks to wondering whether the information will make any impression at all. In some ways, this isn't surprising—after all, many people with high cholesterol or high blood pressure don't make lifestyle adjustments, even though the markers have a substantial effect on disease risk. One open question is whether "people perceive the information as more accurate when DNA is being used," says Theresa Marteau, a psychologist at King's College London. Her analysis of published

CREDIT: MARK WEBER/GETTY IMAGES

studies found that rarely is genetic information regarded fatalistically, as predicting inevitable disease. Now she's considering whether there's something about genetics—perhaps its uniqueness or its perceived accuracy—that can help drive healthy behavior, even if it doesn't add much new information about risk.

Marteau is testing this hypothesis in people who have a close relative with Crohn's disease, an inflammatory disorder of the digestive tract, aiming to recruit about 540 individuals. Because of their family histories, the participants have a Crohn's risk of 2% to 6%, compared with 0.1% in the general population. But their risk is also high for another reason: All the people in Marteau's study are smokers, which about doubles their chance of developing the disease.

Marteau wants to know whether adding genetic information to a risk assessment—even if it doesn't dramatically change the actual risk of disease—makes her volunteers more likely to stop smoking. All participants will receive information based on family history, half will also be tested for a gene, *Nod2*, that boosts risk of Crohn's.

For many diseases, researchers are beginning to consider whether certain combinations of gene variants might have a major impact. In January, scientists reported in the *New England Journal of Medicine* that for men with a family history of prostate cancer, five genetic variants together increases risk roughly nine times. This is similar to a *BRCA*-linked risk for breast cancer. About 2% of men harbor four or five of the prostate cancer variants in question, says Jianfeng Xu, a genetic epidemiologist at Wake Forest University School of Medicine in Winston-Salem, North Carolina, who helped lead the study. Xu and his colleagues have formed a company to commercialize the test, which he expects will be sold to doctors starting late this spring.

Although a ninefold boost in risk for prostate cancer is substantial, Xu still worries that the measure is crude. Only about 15% to 20% of men diagnosed with prostate cancer need aggressive treatment, and a much better test would be one anticipating the likelihood of aggressive disease rather than any prostate cancer. Xu is now hunting for more informative variants and hopes to add them to the test when they've been identified.

The public appears hungry for such information. After his paper appeared, Xu received calls and e-mails from people with a family history of prostate cancer wanting to know their genetic risk. "I don't think we've been very good at anticipating that market forces were going to enter into it," says McBride, noting that researchers and physicians have been outraged by the proliferation of companies selling gene tests directly to consumers. DeCODE Genetics in Reykjavik, Iceland, recently began marketing risk tests for type 2 diabetes, atrial fibrillation and stroke, heart attack, and prostate cancer. Its prostate cancer test, released in February, includes the five variants Xu is focusing on along with three others.

Frontier zone

While physicians worry about how to convey the results of genetic tests to patients, some scientists are thinking more about what GWA studies are not turning up. GWA has established a list of 31 genes implicated in Crohn's disease, for example, but there's still a mystery about how the disease is inherited. None of the 31 genes has explained why Ashkenazi Jews—a genetically cohesive group—are disproportionately at risk for Crohn's. There must be a genetic component that "we're just not getting yet," says Judy Cho, head of the Inflammatory Bowel Disease Center at Yale University. Cho is beginning studies of copy number variation: duplication or loss of DNA stretches that may contribute to disease but aren't picked up in GWA scans.

Inherited risk? A Canadian group is looking for DNA that distinguishes people with a normal heart (*top*) from those with an occluded artery (*bottom*).

CREDIT: DR. MANNO, JUNAZ/UNIVERSITY OF OTTAWA HEART INSTITUTE

In type 2 diabetes, notwithstanding some widely hailed GWA discoveries, "the proportion of heritability that we're picking up is relatively small, and that places serious limits on how good these are as individualized predictions in diagnostics," says Mark McCarthy, an endocrinologist at the University of Oxford in the U.K. "Despite all the giddy excitement, we're only capturing a very small proportion of what's out there." He suggests looking more closely at rare mutations, which, like copy number, are territory that GWA is not designed to capture.

GWA studies could also be skewed by the populations they cover. So far, most have been done on people of European origin, and it's not clear how well they'll translate to other groups. In heart disease, a DNA stretch called 9p21 that falls between genes has been replicated in four large GWA studies, but it's not standing out yet in African Americans. "We're still a little bit unclear about that," says Ruth McPherson, an expert in cardiovascular risk prevention at the University of Ottawa Heart Institute in Canada and one of those who first discovered 9p21. She says that she and her colleagues are "trying to understand if it's not a risk factor for disease among blacks." Complicating matters are the logistical challenges: Many patients in GWA studies of heart disease, says McPherson, are not defined based on coronary angiography, x-ray imaging of the arteries of the heart. Finding people who have undergone angiography to serve as controls is another "difficult problem," she says. Differences among populations and study designs suggest that failure to replicate a finding doesn't necessarily mean it was wrong, she and others say.

That argument isn't persuasive to David Cox, chief scientific officer of the company Perlegen Sciences in Mountain View, California. "You can come up with all sorts of reasons why [a GWA study] didn't replicate, but if you want to use this for any kind of prediction, it has to show up over and over again," he says. Cox worries that scientists are engaged in a love affair with GWA to the neglect of other strategies that can dissect disease. "It's like a huge stampede," he says.

One limitation of the technology, Cox notes, is that it likely won't pick up genetic variation that correlates with a person's response to medical treatments because the number of participants in a study would need to be enormous. Cox is working on this problem; his approach relies on rare gene variants to detect an effect that may be more potent than that conferred by common ones. Cox also questions how often GWA will lead researchers to new drug targets, the possibility has generated much hope and may yet pan out.

All these lingering questions are best solved by one strategy, "Find more of the genes," says McCarthy. "Think of how little of the territory we've covered." That solution—keep up the hunt—is drawing broad support. **–JENNIFER COUZIN**

Cooking Up the Solar System From the Right Ingredients

Cosmochemist Kevin McKeegan had good news for his teammates on the ill-fated Genesis mission. On the way back from deep space, their capsule carrying wispy samples of the solar wind had "hard landed" in the Utah desert in 2004 after its parachute failed to open (*Science*, 17 September 2004, p. 1689). But with luck, a dash of cleaning solvent, and an 8-ton analytical instrument built specially for the mission, "the experiment has worked," McKeegan reported. He and colleagues have successfully measured the oxygen isotopic composition of the sun and therefore of the solar system's starting material. It turns out yet again that—isotopically, at least—Earth is not at the center of things.

Phoenix-like. The crashed Genesis capsule still yielded samples that revealed the sun's true isotopic composition.

Cosmochemists had mapped out the varying isotopic composition of oxygen—the major element of rocky planetary bodies—in Earth, the moon, Mars, and a variety of meteorites. The closer two bodies plotted on the isotope map, the more closely they were related. But "the trouble is we don't know how to read the map," said McKeegan, because no one knew the isotopic composition of the sun, which should match that of the matter from which the solar system formed. For many years, most researchers assumed that the sun's oxygen-isotope composition was the same as Earth's. They thought the extra dose of oxygen's lighter isotope, oxygen-16, in some components of meteorites must have been injected into the planetary mix by a nearby supernova.

To test that supposition, Genesis swept up solar-wind atomic particles with a variety of collecting surfaces, stored them in a return capsule, and headed back to Earth. As it happened, the oxygen collector was suspended on radial arms, which broke its fall on impact and kept its four collecting plates largely intact. The rest of the Genesis collecting surfaces shattered into thousands of pieces. For the analysis, McKeegan and his colleagues used an instrument dubbed MegaSIMS, located in his lab at the University of California, Los Angeles. MegaSIMS could gently clean the collector surface with an ion beam and then bore into a 100-micrometer spot to drive off a whiff of embedded solar-wind oxygen atoms.

With the first results in, McKeegan announced that the meteorites don't have the odd isotopes—Earth does. Somehow the composition of the stuff going into Earth was changed, whereas some parts of meteorites got the sunlike stuff. That argues against a supernova origin, McKeegan said.

Cosmochemist Robert Clayton of the University of Chicago in Illinois has been on both sides of the arguments. "I defended [an external source] for 30 years against all sorts of criticism," he says. "I must say I was pretty stubborn about it." But by 2002, with additional support for a supernova source failing to materialize, Clayton proposed instead that the blazing ultraviolet radiation of the nascent sun skewed the isotopic composition of solar system ingredients. In the process, the irradiation would have vastly altered the way oxygen—which was initially stored largely in carbon monoxide—became incorporated into rocky dust in exposed parts of the disk of gas and dust that would give rise to the planets. Now, McKeegan says, this photochemical scenario offers the best explanation for the oxygen isotopic map of the solar system.

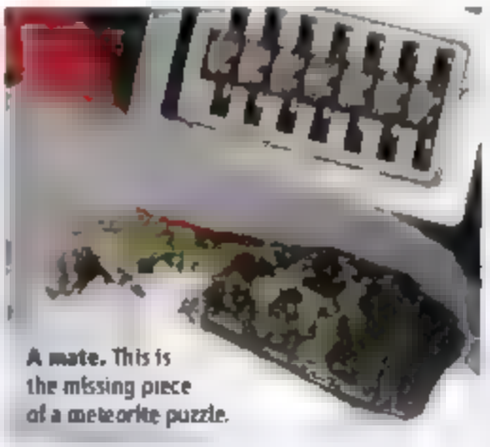
New Piece of the Solar System Puzzle Fits In

With 45,000 bits of the asteroid belt collected from Antarctica alone, you'd think there wouldn't be much new under the sun for meteoriticists to ponder. But in a special session at the meeting, researchers gathered to consider two rocks—broken pieces of a meteorite that fell to the Antarctic ice millennia ago—that had defied classification.

The first chips off Venus, perhaps? No such luck, but the meteorites did turn out to be a missing piece of the puzzle that meteoriticists have been assembling for a century. How did the first bits of the rocky solar system come together and evolve before forming Earth and the other rocky planets? Through melting of primordial stuff, in this case, but many missing pieces of the puzzle remain.

American meteorite hunters collected GRA 06128 and GRA 06129, as they are properly called, from the ice near Graves Nunataks in 2006. But preliminary chemical and mineralogical analyses failed to place the Graves meteorites in any of the more than 60 groups, subgroups, grouplets, and clans meteoriticists use to classify new finds. So samples went out last September for intensive analysis to five multi-institutional consortia that reported their results at the session.

First off, the Graves meteorites aren't from Venus. Their age—more than 4.5 billion years, as reported by three groups—dates back to shortly after the planets and asteroids formed; on average, the venusian surface formed 4 billion years later. The consortia's analyses and inferences tell a consistent story of another origin, says meteoriticist Ralph Harvey of Case Western Reserve University in Cleveland, Ohio, who was not a consortium member. Not far beneath the surface of a



A mate. This is the missing piece of a meteorite puzzle.

CREDIT (TOP TO BOTTOM): USAF 30TH RANGE SQD./ AMSMET

Snapshots From the Meeting >>

Remnant martian ice. Scientists have long known about water ice piled in great sheets at the martian poles and locked in soils around the poles, but at the meeting, researchers reported finding kilometer-thick ice in mid-latitudes as well. Jeffrey Plaut of the Jet Propulsion Laboratory in Pasadena, California, and team members operating the Shallow Radar instrument on the Mars Reconnaissance Orbiter found nearly pure ice within ridges and glacial-looking features that geologists had suspected contained at least some ice mixed with rock. Geologist James Head III of Brown University and colleagues reported that such debris-covered ice features are likely the remnants of more extensive ice sheets twice as thick. They would have grown 100 million years ago or so, when martian climate favored snowfall at lower latitudes.

Too much splat. The 15-meter-diameter crater blasted out of an arroyo in Peru near Lake Titicaca last 15 September shouldn't be there. Small, weak meteorites blazing into the atmosphere at 50,000 kilometers per hour are supposed to break up, slow down, and simply fall to the ground. But impact specialist Peter Schultz of Brown University and his colleagues reported finding bits of weak, stony meteorite around the Peruvian crater along with abundant signs of a hypervelocity impact. How a fragile object about a meter across could come in so fast puzzled other impact specialists at the meeting, but Schultz has an idea. Laboratory experiments he ran 20 years ago suggested to him that under the right conditions, a meteor's cone-shaped shock wave could corral fragments clustered together within the cone so that they would not slow down much. They could then punch into the surface as the Peruvian impactor seems to have done. **–R.A.K.**

What Was a 'Wet and Warm' Early Mars Really Like?

"It's all a mush in my mind," said Michael Carr, talking about early Mars. That's quite a concession from one of the world's most experienced Mars geologists. After 10 hours of discussion on the state of early Mars, "it's frustrating," he said when asked to sum up a small premeeting workshop.* "Despite the beautiful data we've seen, we're no closer to understanding what [early Mars] climate was. And if it was warm and wet, what caused it to be warm and wet?"

Carr, who has been studying Mars since the early 1970s at the U.S. Geological Survey in Menlo Park, California, wrote the book on Mars. Actually, he wrote two of them: *Water on Mars* and *The Surface of Mars*. But that wasn't enough to sort out 40 years of data, much of it still coming from two rovers and two orbiters. The workshop focused on the first period of martian geologic history, called the Noachian—when water flowed on the surface, at least at times—and the transition into the Hesperian, a colder, drier time preceding the bone-dry deep-freeze of the past 3 billion years.

This Noachian-Hesperian transition "is a complicated story," Carr said. "A lot of things happened." Climate got colder and drier. Noachian weathering made clays, then Hesperian weathering made acidic sulfates. The martian magnetic field died, exposing the atmosphere—and the planet's water—to being stripped away into space by the solar wind. The great Tharsis volcanoes quieted down, cutting off their gushing greenhouse gases. And the rain of huge meteorites—each of which may have triggered a geologic moment of warmth and wetness—slowed to a drizzle.

Linking one or more of these potential climate change triggers to the great chilling and drying is what's frustrating Carr. "As a geologist, I'm interested in timing," he said. If the shutdown of Tharsis, for example, closely coincided with the climate shift, he could at least tentatively infer that volcanic gases had been sustaining a strong greenhouse and some sort of warm and wet Mars. But that kind of timing has remained beyond reach for 40 years, he noted.

No one disputed Carr's take on the whys and wherefores of early Mars climate, but others at the workshop were more upbeat about the future. Recent developments in Mars studies are "very exciting," said workshop organizer James Head III of Brown University in response to Carr's remarks. Head's 4 decades in planetary geology have taken him—intellectually—from Mercury to the icy moons of Jupiter, and he is optimistic. New information "always raises new questions," he said. "I'm stunned by the progress" on Mars in recent years. His Brown colleague, John Mustard, took the middle ground. "We certainly aren't finished with the observations," he said. "Hopefully, that won't be just more mysteries." **–RICHARD A. KERR**

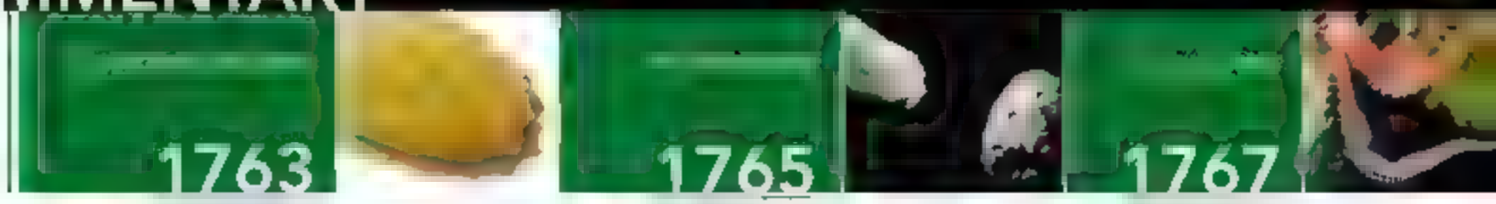
* Brown-Vernadsky Microsymposium 47, 8–9 March, Houston, Texas (www.planetary.brown.edu/html_pages/micro47.htm)

still-hot, sizable planetesimal—perhaps between 100 and several hundred kilometers across—some of the original, unaltered building material of the solar system melted a bit but not completely. The resulting chemically distinct liquid rock left its unmelted residue behind and rose toward the top of a magma chamber, where it slowly cooled, crystallized, and altered under the lingering heat. Eons later, impacts liberated the frozen melt rock, which made its way to Earth.

At the meeting, speakers for the consortium noted that these may not be the first rocks from the Graves "parent body" to be recognized on Earth. Researchers had grouped a half-dozen meteorites together because they were all residues that had some melt extracted from them, and they had similar chemistries, mineralogies, and oxygen isotope compositions. But the melt rock that must have been derived from these so-called brachinite meteorites was now here to be found in the world's meteorite collections. The oxygen isotope composition of the Graves meteorites is a good match to the brachinite group, as is their ratio of iron to manganese. So the missing piece of the brachinite puzzle may have been found, says Harvey.

Even if the Graves meteorites and the brachinites are related, there's more work to be done on the early solar system puzzle. "We just don't know the variety of things out there," says Harvey, who runs the U.S. Antarctic search program. "A lot of us are still butterfly and shell collecting. We still spend most of our time describing and cataloging rather than fitting pieces together, because there are gaps." Forty-five thousand, it seems, is not enough.

CREDIT: NASA/JPL/MALIN SPACE SCIENCE SYSTEMS



LETTERS

edited by Jennifer Sills

The Last Inventor of the Telephone

SETH SHULMAN'S BOOK "THE TELEPHONE GAMBIT" AND ITS FINE REVIEW BY D. L. MORTON Jr. (29 February, p. 1188) focus on the 1876 dispute between Alexander Graham Bell and Elisha Gray over who invented the telephone. In fact, neither of them was first. Interestingly, different views on this topic prevail in different nations. French accounts tend to emphasize Charles Bourseul's theoretical underpinnings of the phone (1854). Many Italians, meanwhile, consider Antonio Meucci to be the real inventor—his phone apparently was operational in 1857 (acknowledged by a 2002 bill of the U.S. House of Representatives). Germans frequently cite the 1860 electric telephone by Philipp Reis. Compared to all these pioneers, Gray and Bell came rather late. Bell is championed in his home country, Scotland, his adopted home, Canada, and the United States (he became a U.S. citizen 6 years after filing his patent). Unlike his predecessors, however, Bell was able to create a successful phone company, and he thus acquired financial and public relations resources that helped to widely promote his own view of who invented the phone.

What can we learn from this? When the time is ripe for an invention, it tends to be pursued and developed in various places until someone manages to make a public breakthrough. At least in popular culture, much of the credit is bestowed upon the last contributor, even when the essential original insights came from others. As they say, Columbus did not become famous because he was the first to discover America, but because he was the last.

JÜRGEN SCHMIDHUBER

Istituto Dalle Molle di Studi sull'Intelligenza Artificiale, 6928 Manno-Lugano, Switzerland, and TU München, Fakultät für Informatik, Garching bei München, Germany

Thinking Outside the Reef

A RECENT REVIEW BY O. HOEGH-GULDBERG *et al.* ("Coral reefs under rapid climate change and ocean acidification," 14 December 2007, p. 1737) warns that ocean acidification will compromise carbonate accretion, with accelerating functional collapse of coral reef ecosystems worldwide if atmospheric CO_2 rises above 500 parts per million. However, readers were not offered a way to prevent the demise of coral reefs.

It is imperative that resource managers and researchers promote ecologically sustainable development (ESD) on a global scale. We call on them to lead by example by investing in energy-efficient practices and sustainable forestry. The most practical means toward the goal of carbon neutrality involve buildings and forests, which are worth 22 and 14%, respectively, of necessary global CO_2 control (*1*).

Energy use in buildings offers the largest share of cost-effective opportunities for CO_2 mitigation, with ESD strategies such as daylighting, improved ventilation, cool roofs, shading, and insulation (*1*). Sustainable forestry in coastal zones gives an added benefit to coral reefs by buffering the seaward run-off of nutrients and sediments (*2*). To overcome business-as-usual "paradigm blindness," resource managers, research institutions, and tourist operators who are dedicated to ecosystems at risk should use benchmarking to expose new methods, ideas, and tools (*3*, *4*). We recommend systematic auditing and target-setting of CO_2 emissions and sinks. To save coral reefs, stakeholders must make management of CO_2 part of their core business, with the understanding that control is impossible without monitoring. Institutions need to disseminate ESD principles to the wider community—for example, by adding insulation to existing buildings and planting trees in riparian zones—to augment direct coral reef management actions such as promotion of herbivorous parrotfish.

ERIC L. PETERSON,[1] MARIA BEGER,[2] ZOE T. RICHARDS[3]

[1]Department of Architectural, Civil, and Mechanical Engineering, Victoria University Melbourne MC, Victoria 3032, Australia. [2]The Ecology Centre and The Commonwealth Research Facility for Applied Environmental Decision Analysis, University of Queensland, St. Lucia, QLD 4072, Australia. [3]Australian Research Council Centre of Excellence for Coral Reef Studies, and School of Marine and Tropical Biology, James Cook University, Townsville, QLD 4811, Australia.

References

1. IPCC 2007. *Climate Change 2007: Mitigation of Climate Change. Contribution of Working Group III to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change*. B. Metz, O. R. Davidson, P. R. Bosch, R. Dave, L. A. Meyer, Eds. (Cambridge Univ. Press, Cambridge, 2007).
2. K. E. Fabricius, *Mar. Pollut. Bull.* **50**, 125 (2005).
3. Office of Government Commerce, Gary Case. *2007 Continual Service Improvement* (The Stationery Office, Norwich, UK, 2007).
4. T. Markus, J. A. Paravantis, *Energy Build.* **39**, 404 (2007).

CREDIT: JOE SUTLIFF

Putting Ant-Acacia Mutualisms to the Fire

VARIOUS COMPLEX TRADE-OFFS DETERMINE the vitality and survival of acacias in dynamic savannas. Recently, T. M. Palmer *et al.* ("Breakdown of an ant-plant mutualism follows the loss of large herbivores from an African savanna," Reports, 11 January, p. 192) illustrated how exclusion of large herbivores changed ant-plant mutualistic

relationships on *Acacia drepanolobium* at a site in upland Kenya, emphasizing the dominant role of mammalian browsing in this system. However, when browsing on acacias decreases, so typically does trampling and grazing around the trees, increasing the risk of fires. These factors were not considered in their interpretation.

At Mkwaja Ranch on the Tanzanian coast, homogeneous woodlands of tall, pole-like *Acacia zanzibarica* trees—a myrmecophyte equivalent of *A. drepanolobium*—had established between 1954 and 2000 under intensive cattle grazing and in the virtual absence of browsing by wildlife (*1*–*3*). In 2002, the ant *Crematogaster sjostedti* occupied 99% of the trees with ant colonies (*2*, *4*). After ranching had ceased in 2000, a continuous grass layer was established, and hot fires defoliated some woodlands up to about 10 m above ground in early 2003, resulting in 23% mortality of mostly smaller trees (*2*). Tunneling by wood-boring beetles—common in dead wood—may have increased mortality of smaller already weakened trees, but tunneling was also found in apparently healthy trees, allowing ants to retire during fires (*2*).

Three weeks after fires, an unseasonal foliage flush was observed whereby damaged trees produced more foliage than unaffected trees (*2*). Such costly investment by trees may serve to reboost resident *C. sjostedti* populations surviving within the tunnels, via new nectar-secreting glands on leaf petioles. A complete loss of ant symbionts—invariably leading to high levels of insect herbivory on nutritious foliage (*5*–*7*)—may, however, have fatal consequences for myrmecophyte acacias (*8*). Fire and ant-plant-insect interaction deserve more scientific attention as potentially important evolutionary agents driving ant-acacia mutualisms in Africa.

ROLAND COCHARD[1]* AND DONAT AGOSTI[2,3]

[1]Institute of Integrative Biology, Swiss Federal Institute of Technology, 8092 Zurich, Switzerland. [2]Division of Invertebrate Zoology, American Museum of Natural History, [redacted] Burgergemeinde Bern, 3005 Bern, [redacted]

*To whom correspondence should be addressed. E-mail roland.cochard@env.ethz.ch

References and Notes

1. M. W. Tobler, R. Cochard, P. J. Edwards, *J. Appl. Ecol.* **40**, 430 (2003).
2. R. Cochard, *Patterns and dynamics of secondary Acacia zanzibarica woodlands at Mkwaja Ranch, coastal Tanzania* (Swiss Federal Institute of Technology, Zurich, 2004).
3. A. C. Treydte, P. J. Edwards, W. Suter, *Afr. J. Ecol.* **43**, 302 (2005).
4. *Crematogaster sjostedti* Mayr has now been identified.
5. M. L. Stanton, T. P. Young, *Nat. Hist.* **11**, 28 (1999).
6. D. H. Janzen, *Evolution* **20**, 249 (1966).
7. M. T. Shaw, F. Keesing, R. S. Ostfeld, *Oikos* **98**, 385 (2002).
8. D. H. Janzen, *Ecology* **48**, 26 (1967).

Response

COCHARD AND AGOSTI ARE CORRECT THAT interactions between herbivory and fire may influence the dynamics of ant-plant symbioses in African savannas. Their observations in a lowland site in Tanzania indicate that these influences are likely to be complex and site-dependent. In their study system, loss of grazers produced increased ground vegetation cover, increased fire intensity, and possibly increased dominance of the cavity-dwelling plant ant *Crematogaster sjostedti*. The precise pathway of causation suggested by this scenario differs from that operating in upland Kenya, but their observations do not alter our study's conclusion that browsing mammals can maintain ant-plant mutualisms in ecological time.

Our results cannot be explained by the effects of herbivore exclusion on fire regime. Fire is actively suppressed throughout much of the commercial rangeland that encom-

passes our study sites (i.e., natural and accidental fires are extinguished by fire crews, and the spread of fires is contained by the placement of fire breaks throughout the landscape) (*1*). Several experimental burns have suggested that more frequent fires would not favor *C. sjostedti* over other species in the acacia-ant guild. Similar to the scenario described by Cochard and Agosti, our fire experiments reveal disproportionate postfire survival of *C. sjostedti* colonies sheltered within large trees. However, *C. mimosae* and *C. nigriceps* also display high colony survivorship due to effective evacuation behaviors. Colonies of both species evacuate all workers, brood, and winged reproductives to insulated cracks in the soil within approximately 45 minutes of smoke's reaching the tree. The fourth ant species in our system, *Tetraponera penzigi*, which inhabits mostly small trees, suffers nearly 100% colony mortality during a fire but is a stronger colonist of unoccupied trees than *Crematogaster* (*2*) and may therefore reassert itself in fire-thinned landscapes. Although herbivory and fire regimes frequently interact in African savannas, the processes observed by Cochard and Agosti would likely reinforce the dynamics we documented in the absence of fire.

Cochard and Agosti's assertion that decreases in browsing intensity are typically correlated with decreases in grazing intensity may not hold true in all cases. In our study system, the density of ground vegetation in the vicinity of adult *A. drepanolobium* is driven largely by cattle, and the removal of native browsers may or may not be associated with increased herbaceous cover. Our experimental design reflected this situation (both treatments contained plots with and without cattle, which therefore differed in understory cover). Thus, whatever additional complexity fire might add to the equation, our findings do not hinge on a correlation between grazing and browsing intensity. We also note that postfire regrowth, bringing "new nectar-secreting glands on leaf petioles," would be unlikely to bolster *C. sjostedti* populations in our study system, where *C. sjostedti* does not intensively use nectaries (*3*).

Despite the differences in environmental context between upland Kenya and lowland Tanzania, the series of events outlined by Cochard and Agosti produce a net outcome qualitatively similar to that described in our study. In both cases, the exclusion of large mammals (browsers in our study, grazers in theirs) favors community dominance of the cavity-nesting ant *C. sjostedti*, although this occurs via different mechanistic pathways.

TODD M. PALMER,[1] MAUREEN L. STANTON,[2] TRUMAN P. YOUNG,[2] JACOB R. GOHEEN,[3] ROBERT M. PRINGLE,[4] RICHARD KARBAN[2]

[1]University of Florida, Gainesville, FL 32611, USA. [2]University of California, Davis, CA 95616, USA. [3]University of British Columbia, Vancouver, BC V6T 1Z4, Canada. [4]Stanford University, Stanford, CA 94305, USA.

References

1. O. J. Augustine, S. J. McNaughton, *J. Appl. Ecol.* **41**, 45 [illegible]
2. M. L. Stanton, T. M. Palmer, T. P. Young, *Ecol. Monogr.* **72**, 347 (2002).
3. T. M. Palmer *et al.*, *Science* **319**, 192 (2008).

Letters to the Editor

Letters (~300 words) discuss material published in *Science* in the previous 3 months or issues of general interest. They can be submitted through the Web (www.submit2science.org) or by regular mail (1200 New York Ave., NW, Washington, DC 20005, USA). Letters are not acknowledged upon receipt, nor are authors generally consulted before [illegible]. Letters are subject to editing for clarity and space.

SCIENCE AND THE ARTS

Another Approach to Consilience

Jay Labinger

> When we have unified enough certain knowledge, we will understand who we are and why we are here.
>
> –E. O. Wilson, *Consilience*

> When we have found all the mysteries and lost all the meaning, we will be alone, on an empty shore.
>
> –Tom Stoppard, *Arcadia*

In *Consilience* (*1*), Wilson called for the unification of the scientific and humanistic spheres of learning—C. P. Snow's famous "two cultures." He proposed that the way to achieve it is "to view the boundary between the scientific and literary cultures not as a territorial line but as a broad and mostly unexplored terrain awaiting cooperative entry from both sides." It would be hard to disagree with that sentiment, but Wilson's claim that "the only way either to establish or to refute consilience is by methods developed in the natural sciences" seems to have much more to do with colonization than alliance. Indeed, the substantial amount of work toward consilience consists primarily of examining some aspect of the humanities from a scientific perspective. Some notable recent book-length examples include an anthology of Darwinian literary criticism (*2*) as well as explorations of the implications of neuroscience for the creation of, and response to, the visual arts (*3*) and music (*4*).

A madeleine: "the trigger for Proust's epiphany."

In contrast, Jonah Lehrer's *Proust Was a Neuroscientist* and David Edwards's *Artscience: Creativity in the Post-Google Generation* come from a different direction. Although they do not approach the issue in quite the same way, both can be taken as arguing that scientific knowledge may equally well be informed by humanistic study. And both authors offer a good deal of eloquent, if not always completely satisfying, support for their stance.

Lehrer (who writes the blog Frontal Cortex, http://scienceblogs.com/cortex) examines five authors plus one representative each from the visual, musical, and culinary arts. He contends that their work reveals important information about how the brain functions: not mere analogy or metaphoric description, but real truths that are no less valid in their realm than those obtained from more reductionist scientific methods. These include purported demonstrations that Walt Whitman's poetry evokes the intimate interconnections between mind and body, refuting "Descartes' error" more than a century ahead of Antonio Damasio (*5*); that Marcel Proust, in his massive, introspective *A la Recherche du Temps Perdu*, constructed a model of the workings of memory that seems in many ways fully consistent with the findings of recent study, nearly down to the molecular level; and that Auguste Escoffier's methods of cookery presciently anticipated the subsequent discovery of umami, the fifth taste.

I found a number of his arguments rather compelling. On the other hand, a few cases are less convincing, especially when Lehrer allows an apparent predilection for hyperbole and overblown rhetoric to trump clear and reasoned explanation. More seriously, his presentation of the scientific content is too frequently imprecise. For example, the discussion of the chemistry of glutamate is quite confused. There are even some outright elementary blunders, as when thymidine and the other components of DNA are misidentified as amino acids. These defects diminish the overall impact of the book. Worse, they furnish ammunition to those critics (whom Lehrer explicitly takes on, in his "Coda") who dismiss Lehrer's knowledge claims as nonrigorous and hence irrelevant.

Edwards's book has more of the character of a practical "how-to" manual. His credo—one that will probably strike most readers as much less contentious than Lehrer's—is that people who can transcend the cultural and intellectual boundaries between science and the arts thereby become more creative and productive. Edwards (a professor of bioengineering at Harvard) calls this pathway "artscience" (Lehrer uses "fourth culture" for much the same thing), and a large part of the book consists of introducing individuals who follow it. Several of these stories do inspire, including the musician who re-educates herself as an engineer and invents a new method of composition, the chemical engineer whose artistic background inspires a theory of fluid mixing, and Edwards's own experience in establishing a center, Le Laboratoire (*6*), to promote "idea translation." As in Lehrer's book, not all the examples are equally persuasive. (Also, at the risk of sounding too cynical, I can't help wondering about the counter-stories one might tell of those who spread themselves too thinly and accomplish little or nothing.) Furthermore, when Edwards leaves the life stories for more general discussion, he occasionally descends into somewhat painfulness exhortations for collaborating and thinking outside the box.

Despite these reservations, I found both books entertaining and generally worthwhile. At a minimum, they provide useful support for the point that (I believe) Stoppard was trying to make in the quote above: trying to get to Wilson's promised land by means of only the navigational tools provided by science is likely to lead us astray. True consilience will require the ability of science to solve mysteries and the ability of the arts and humanities to produce meaning, with equal respect from and for both sides.

The reviewer is at the Beckman Institute, California Institute of Technology, 1200 East California Boulevard, Pasadena, CA 91125, USA. E-mail: jal@its.caltech.edu

References and Notes

1. E. O. Wilson, *Consilience: The Unity of Knowledge* (Knopf, New York, 1998); reviewed by J. Dupré, *Science* **280**, 1395 (1998).
2. J. Gottschall and D. S. Wilson, Eds., *The Literary Animal: Evolution and the Nature of Narrative* (Northwestern Univ. Press, Evanston, IL, 2005); reviewed by H. Fromm, *Science* **311**, 612 (2006).
3. B. M. Stafford, *Echo Objects: The Cognitive Work of Images* (Univ. of Chicago Press, Chicago, 2007); reviewed by C. A. Jones, *Science* **319**, 35 (2008).
4. D. J. Levitin, *This Is Your Brain on Music: The Science of a Human Obsession* (Dutton, New York, 2006).
5. A. Damasio, *Descartes' Error: Emotion, Reason, and the Human Brain* (Dutton, New York, 1994).
6. www.laboratoire.org; see review by L. Whiteley, *Science* **318**, 1871 (2007).

10.1126/science.1155967

HISTORY OF SCIENCE

Explaining the Cosmos to the People

Philip J. Pauly

"I sometimes think" Charles Darwin wrote in an 1865 letter to Thomas Henry Huxley, "that general and popular Treatises are almost as important for the progress of science as original work" (*1*). Publishers would have killed to use that statement as a testimonial. In *Victorian Popularizers of Science*, Bernard Lightman sketches the landscape of reading in Britain during the middle third of the 19th century, explores the extent to which readers were interested in nature and science, and outlines the efforts of publishers who used steam presses and mass marketing to meet these demands and sell more products.

Lightman (a historian of science at York University, Toronto) presents extended biographical sketches of about 30 popularizers who wrote between the 1830s and the 1910s, giving greatest emphasis to the period from 1860 to 1890. His book will be the basic resource for scholars interested in understanding the background of the thousands of popular science works that sit patiently on university library shelves and are sometimes, due largely to Dover Publications, still in print. It includes a path-breaking recovery of the lives, interests, and limitations faced by female nature writers such as Jane Loudon, Arabella Buckley, and many less-familiar women. Popularizers' output encompassed introductions directed at children, the elucidation of the principles underlying ordinary household phenomena, practical guides for identifying natural objects, and holistic perspectives on the countryside and its inhabitants.

There was a grander theme, however. Beginning with the anonymously published *Vestiges of the Natural History of Creation* (*2*), popularizers propounded the "evolutionary epic." Encouraged by the writings of Herbert Spencer and Darwin, materialists such as Edward Clodd and Grant Allen emphasized competition, selection, and their culmination in human evolution. Astronomical writers such as Richard Proctor and Agnes Clerke, by contrast, pointed to signs of ancient extraterrestrial life and to the magnificence of galaxies in order to convey the likelihood of universal design. Lightman concludes that, whatever answers they advocated, Victorian popularizers provided the public a large vision of the scope and philosophical import of evolution. They shaped public interest up to the present, and Carl Sagan, E. O. Wilson, Stephen Jay Gould, and NASA are among their heirs.

Unfortunately for the general reader in 2008, Lightman does not emulate his subjects' fundamental insight that a clear narrative line interspersed with synoptic views of the intellectual landscape is the key to successful writing. His evidential base is his series of biographies. He orders these, however, neither chronologically nor thematically. Post-Darwinian Anglican naturalists (such as Charles Kingsley) are followed by mostly pre-Darwinian female journeyman writers and then by later showmen. He suddenly introduces the evolutionary epic of the 1840s halfway through the text. His discussion of scientific practitioners as popular writers is limited to two figures, Huxley and the astronomer Robert Ball.

What is missing is the middle landscape of gentlemen, naturalists, and intellectuals, including Sir John Lubbock, George Romanes, and others, who had secure finances and who wrote because they had original things to say on many subjects. This point matters because it clarifies the peculiar cultural status of the evolutionary epic. Lightman emphasizes the polemics between nature writers with often weak backgrounds and specialized practitioners such as Huxley and *Nature* editor Norman Lockyer. If we focus instead on the disciplinarily ill-defined but highly visible networks of philosophers, psychologists, anthropologists, archaeologists, and historians, and explore the ways they created and fed their audiences, we can grasp the space occupied by a specifically "historical" science—for which the evolutionary epic, theistic or not, supplied the framework.

Stephen Gould's *Wonderful Life* (*3*) brought this issue to the fore. He emphasized that historical narratives—whether about humans, the extinction of the dinosaurs, or the more obscure period of the Burgess Shale—were real knowledge. Martin Rudwick's *Bursting the Limits of Time* (*4*) delineates the deeper historical background of this approach. Scientists reluctant to embrace the position that historical accounts are real knowledge hobble their ability to speak persuasively about evolution. Particularly within the American context, where creationists have distorted scientific common sense, dancing with the meanings of theory is a less effective public posture than the four-square presentation of a great historical fact.

Another place for life. Richard Proctor depicted Earth-like continents and seas on Mars (*5*).

CREDIT: FROM *OTHER WORLDS THAN OURS* (S/C COURTESY UNIVERSITY OF CHICAGO PRESS)

The reviewer is at the Department of History, Rutgers University, 16 Seminary Place, New Brunswick, NJ 08901–1108, USA. E-mail: pauly@rci.rutgers.edu

References

1. F. Burkhardt *et al.*, Eds., *The Correspondence of Charles Darwin, Vol. 13, 1865* (Cambridge Univ. Press, Cambridge, 2002).
2. [R. Chambers], *Vestiges of the Natural History of Creation* (John Churchill, London, 1844).
3. S. J. Gould, *Wonderful Life: The Burgess Shale and the Meaning of History* (Norton, New York, 1989).
4. M. Rudwick, *Bursting the Limits of Time: The Reconstruction of Geohistory in the Age of Revolution* (Univ. of Chicago Press, Chicago, 2005); reviewed by N. Oreskes, *Science* **314**, 596 (2006).
5. R. Proctor, *Other Worlds Than Ours* (Longmans, Green, London, 1870).

10.1126/science.1156188

PLANETARY SCIENCE

The Planet Debate Continues

The IAU decisions a year ago to define "planet" narrowly and to demote Pluto have not been universally accepted.

Mark V. Sykes

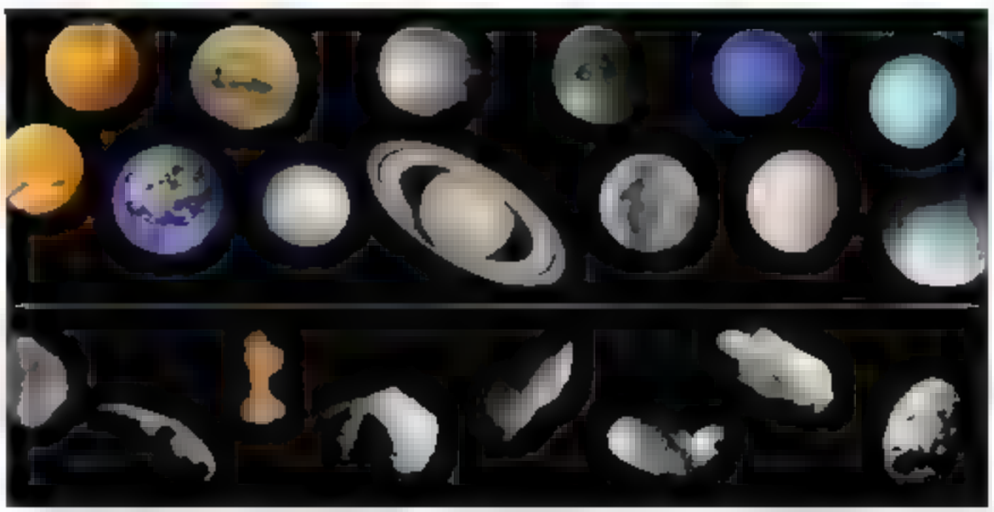
It has been over a year and a half since the International Astronomical Union (IAU) decided to officially define the term "planet" in a way that would demote Pluto (*1*). The result was a public brouhaha and the unfortunate impression that science is done by vote in a conference hall. Hundreds of planetary scientists signed a protest petition, rejecting the IAU definition (*2*).

Science is a never-ending process of refinement and modification, as more information and observations are gathered and theories are proposed and tested. Definitions and classification schemes are a part of that process and are evaluated by their usefulness.

The IAU definition restricts a "planet" to our own solar system and requires that it has "cleared the neighborhood around its orbit" (*1*). This ignores more than 400 objects orbiting other stars that researchers characterize as "planets." Another objection is the lack of clarity about what it means for an orbit to be "cleared." At face value, the large population of Trojan asteroids straddling Jupiter's orbit would seem to preclude the giant from being a planet and could lead to nightmares for high-school teachers everywhere.

The inferred intent of the IAU's definition was to identify those objects that are gravitationally dominant in our solar system. Few people understand an important consequence of this: IAU planets must be more and more massive with increasing distance (semi-major axis) from the Sun. Pluto was originally accepted as a planet, despite its odd orbit, in part because it was erroneously thought to be as massive as Earth. However, under the IAU definition, if Pluto had five times the mass of Earth, it would still not be a planet. In fact, if Mars were just a little farther from the Sun (*3*) (about half an astronomical unit), it would not be a planet. These are not problems for dynamicists, but the definition yields incongruous results for those who study the physical characteristics of planetary bodies.

As more spacecraft are sent to other planets, we gather increasingly detailed information about volcanism, tectonics, atmospheric circulation and chemistry, erosion, fluvial processes, and the potential for life. We benefit from the analysis of this information and the insights by extension to how Earth (the ultimate planet) works. Are there any characteristics shared by all of the bodies exhibiting some or all of these processes? We have now gathered a large database of imagery of other worlds. When we put this imagery together, one thing becomes immediately clear: The objects exhibiting any of these processes are all round. Smaller, irregular-shaped bodies evidence only the history of impacts and volatile loss on their surfaces (see figure, above).



Round solar system bodies exhibit atmospheric and/or geological features related to processes familiar on Earth: (top, first row) Mercury, Mars, Jupiter, Titan, Neptune, and Uranus; (second row) Venus, Earth, Ceres, Saturn, Pluto, Titania, and Triton. Small irregular objects do not show evidence of these processes: (bottom) Tempel 1, Eros, Kleopatra, Gaspra, Ida, Itokawa, Ida, and Wild 2.

CREDIT: NASA

Why is this? Objects become "round" when they are so massive that their gravity crushes them into a shape that is in hydrostatic equilibrium. Heat from formation and the decay of radionuclides increases interior temperatures to the point where differentiation, mantle convection, and other processes occur. In differentiation, heavier denser materials sink to the center while lighter materials rise to the surface. Volatiles that make up oceans and atmospheres arise from this process and are retained if gravity is sufficient. Mantle convection gives rise to tectonic activity and volcanism.

When we identify an object in our own solar system or around another star as being round or sufficiently massive to be round, we expect that terrestrial-analog processes may be found and studied on that object. Thus, a reasonable geophysical definition for the term "planet" may be formed as follows: *A planet is a round object (in hydrostatic equilibrium) orbiting a star.* The number of planets in our own solar system would expand immediately to include Ceres, Pluto and Charon as a double planet, and the distant Eris, for a total of 12. There are likely many more, which leaves open the possibility of future discoveries. The geophysical definition of planet has the advantage of being simple to teach and can be directly related to observations and discoveries made by decades of spacecraft exploration (*4*). It is useful. Further, the IAU "dynamical" planets would be a subset of geophysical planets, which is logical (*5*). The issue continues to be discussed, for example, at an upcoming meeting in August 2008, which is planned to involve teachers (*6*). Whether the IAU ultimately reverses its decision and embraces this larger perspective is not critical to science. While the public is keenly interested in the "outcome," they are best served by being exposed to the debate itself, so that they can see the ongoing process of science.

Planetary Science Institute, 1700 East Fort Lowell, Suite 106, Tucson, AZ 85719, USA. E-mail: sykes@psi.edu

References and Notes

1. G. Schilling, *Science* **313**, 1214 (2006).
2. J. Hogan, *Nature* **442**, 965 (2006).
3. H. Levison, "A hand-waving derivation of planethood" (2006); www.boulder.swri.edu/~hal/planet.
4. Satellites with planet characteristics, such as Saturn's moon Titan, would be distinguished as "planetary satellites," but could also be planets under a definition not requiring that they orbit a star.
5. The IAU defines "dwarf planets" as round objects that have not cleared their orbits, but explicitly says they are not planets, which confuses many.
6. *The Great Planet Debate: Science as Process*; http://gpd.psi.edu.

10.1126/science.1155743

MATERIALS SCIENCE

Multitasking in Tissues and Materials

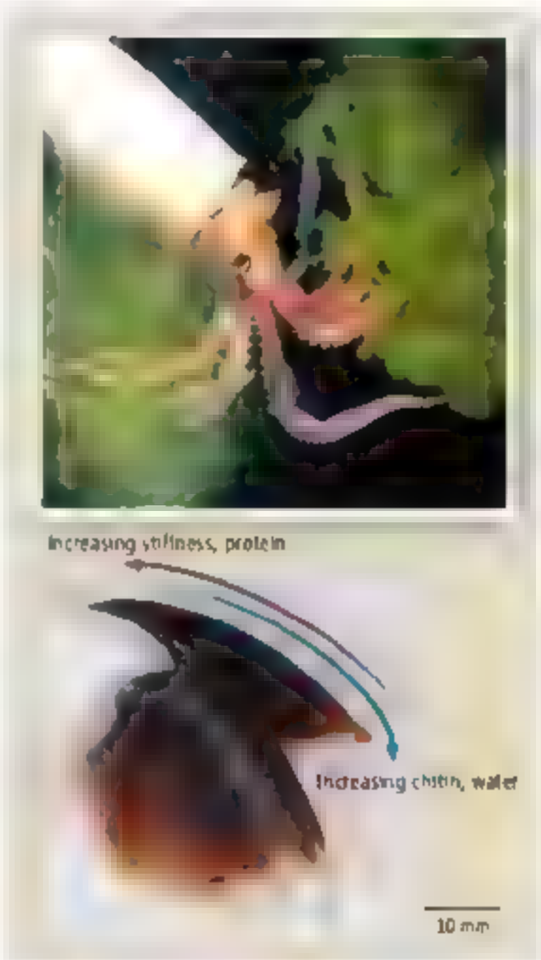
Insights into the role played by a modified amino acid residue in structural biological tissues are helping to develop biomimetic materials.

Phillip B. Messersmith

Many proteins contain posttranslational modifications that are essential for their biological function (*1*). One such modification—the hydroxylation of tyrosine, resulting in the formation of 3,4-dihydroxyphenyl-L-alanine (dopa)—plays a key role in melanin formation and as a biosynthetic precursor to catecholamine neurotransmitters. Dopa is also present in tissues as diverse as mussel byssus, marine tubeworm cements, and the hard mouthparts of marine invertebrates. Studies of dopa-containing tissues have revealed interesting chemical, optical, and mechanical properties, and suggest new pathways toward exploiting dopa and related molecules as biomimetic materials. The report by Miserez *et al.* on the mechanical properties of squid beaks [see page 1816 (*2*)] suggests that dopa is also intimately involved in the formation of mechanical gradients in tissues.

To avoid stress concentrations and tissue damage, nature often employs physical property gradients at interfaces between tissues with vastly different mechanical properties (*3*). For example, at the interface between dentin and enamel in human teeth, a several-fold change in hardness and other mechanical properties occurs smoothly across a width of about 10 μm (*4*). Unlike dentin and enamel, squid beak is not mineralized (*5*). It is nevertheless remarkably stiff, with a modulus of 5 GPa near the beak tip. This value is consistent with the main task of the beak—penetration of the soft tissues of prey (see the figure)—but is several orders of magnitude stiffer than the soft tissue that supports it.

A large, abrupt change in stiffness between beak and soft supporting tissue is minimized by employing a gradient in mechanical properties from tip to base. Nanoindentation experiments performed by Miserez *et al.* revealed a two-orders-of-magnitude gradient in stiffness from tip to base in the squid beak. The beak is primarily composed of chitin, water, and protein enriched in dopa and histidine. Chitin, a water-insoluble polysaccharide also found in insect exoskeletons, is present in squid beak in the form of a nanostructured fibril network. The observed mechanical gradient is closely linked to changes in organic composition (for example, the relative ratios of chitin, water, and protein). Chitin and protein content are inversely related, with the stiffest location—the beak tip—having the lowest chitin and highest protein content.



How squids eat. Squids (top) use a sharp, horny beak (bottom) to cut up food. Miserez *et al.* now show that the beak's stiffness varies by two orders of magnitude from its tip to its base, accomplished by a biochemical gradient of polysaccharide, protein, and water.

Analysis of chemical digests of the beak revealed the existence of dopa-histidine fragments, explaining the dark pigmentation seen in the tissue and suggesting a cross-linking reaction between dopa and histidine. The gradient in dopa-histidine reaction product (reflected by pigmentation) parallels the stiffness gradient, but it remains unclear how the two are correlated and how water contributes to the mechanical properties of the tissue. The pigment of squid beak may have a load-bearing function (*6*).

In addition to the hard tissues of marine invertebrates, dopa has been identified as a key player in the formation of structural biological tissues such as mussel byssal threads and adhesive plaques, and marine tubeworm cement (*7*). The chemical versatility of the catechol side chain of dopa is particularly remarkable. The residue is believed to participate in covalent cross-linking reactions during the solidification of marine adhesive proteins, has a high affinity for transition metals, and strongly adheres to both organic and inorganic surfaces. This is why dopa has captured the interest of scientists and engineers seeking to exploit its unique properties.

Several groups have developed biomimetic adhesives by incorporating dopa or catechol functional groups into synthetic polymers intended for use as industrial, consumer, or medical adhesives (*8*). Recently, a catecholic polymer was employed to fabricate a reversible wet-dry adhesive inspired by mussels and geckos (*9*). Dopa peptides or catechol functional groups have also been used to immobilize synthetic and biological macromolecules onto various substrates (*10*). This strategy has been used to prepare biofouling-resistant polymer coatings (*11*–*13*) and to add biofunctional groups to therapeutic and diagnostic nanoparticles (*14*, *15*). Finally, spontaneous polymerization of dopamine onto surfaces can be used to mediate the attachment of multifunctional coatings onto virtually any material (*16*).

Current synthetic biomimetic materials remain primitive in comparison to their natural counterparts. Our ability to incorporate elements of biological inspiration into the design of synthetic materials will be further enhanced through studies such as that by Miserez *et al.* that advance our understanding of the composition, structure, and processing of complex biological tissues.

Department of Biomedical Engineering, Department of Materials Science and Engineering, and Institute for Bionanotechnology in Medicine, Northwestern University, Evanston, IL 60208, USA. E-mail: philm@northwestern.edu

References

1. C. T. Walsh, G.-L. Sylvie, J. G. J. Gatto, *Angew. Chem. Int. Ed.* **44**, 7342 (2005).
2. A. Miserez, T. Schneberk, C. Sun, F. W. Zok, J. H. Waite, *Science* **319**, 1816 (2008).
3. S. Suresh, *Science* **292**, 2447 (2001).
4. G. W. Marshall Jr., M. Balooch, R. R. Gallagher, S. A. Gansky, S. J. Marshall, *J. Biomed. Mater. Res.* **54**, 87 (2001).
5. C. C. Broomell, R. K. Khan, D. N. Moses, A. Miserez, M. G. Pontin, G. D. Stucky, F. W. Zok, J. H. Waite, *J. R. Soc. Interface* **4**, 19 (2007).
6. D. N. Moses, J. H. Harwell, G. D. Stucky, J. H. Waite, *J. Biol. Chem.* **281**, 34826 (2006).
7. J. Sagert, C. Sun, J. H. Waite, in *Biological Adhesives*, A. M. Smith, J. A. Callow, Eds. (Springer, Berlin, 2006), pp. 125–143.
8. B. Lee, J. L. Dalsin, P. Messersmith, in *Biological Adhesives*, A. M. Smith, J. A. Callow, Eds. (Springer, Berlin, 2006), pp. 257–278.
9. H. Lee, B. Lee, P. B. Messersmith, *Nature* **448**, 338 (2007).
10. H. Lee, N. F. Scherer, P. B. Messersmith, *Proc. Natl. Acad. Sci. U.S.A.* **103**, 12999 (2006).
11. J. L. Dalsin, B.-H. Hu, B. P. Lee, P. B. Messersmith, *J. Am. Chem. Soc.* **125**, 4253 (2003).
12. A. R. Statz, R. J. Meagher, A. E. Barron, P. B. Messersmith, *J. Am. Chem. Soc.* **127**, 7972 (2005).
13. S. Zürcher *et al.*, *J. Am. Chem. Soc.* **128**, 1064 (2006).
14. T. Pauneska *et al.*, *Nat. Mater.* **2**, 343 (2003).
15. C. Xu *et al.*, *J. Am. Chem. Soc.* **126**, 9938 (2004).
16. H. Lee, S. M. Dellatore, W. M. Miller, P. B. Messersmith, *Science* **318**, 426 (2007).

10.1126/science.1155122

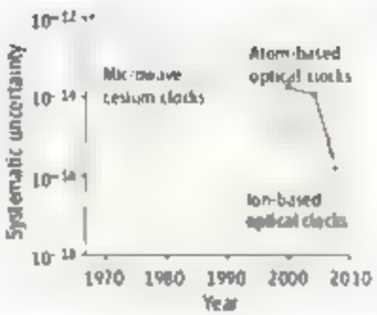
PHYSICS

A Milestone in Time Keeping

Researchers have made atomic clocks so precise that effects of general relativity are on the verge of complicating the concept of keeping time.

Daniel Kleppner

A clockmaker who aspires to hold the record for creating the best clock in the world faces a dilemma. The claim would be unprovable because there would be no absolute reference clock with which to compare it. The only solution would be to build a second clock as good as the first. Two groups have done just that. Rosenband *et al.* report their results on page 1808 of this issue (*1*), and Ludlow *et al.* on page 1805 (*2*). Each group compared two clocks and demonstrated a precision that significantly exceeds that of today's best time standards (see the first figure).



Clocking progress. The systematic uncertainty is the uncertainty due to all known perturbations. Cesium (blue) remains the legal time standard. The ion-based optical clocks (green) operated with Hg^+, Yb^+, Sr^+, or Al^+. The atom-based optical clocks (red) include the pioneering measurement of the hydrogen 1S-2S transition by Hänsch and his colleagues (*7*) (first red point). Other optical clocks operated with Ca and Sr. The final points are for the clocks reported by Rosenband *et al.* and Ludlow *et al.*

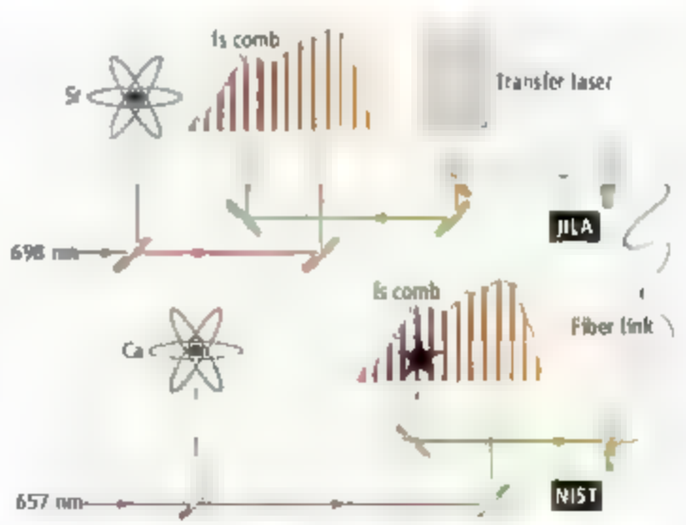
Atomic clocks became a reality in the mid-1950s with the development of the cesium clock. The time-keeping element in this clock is a microwave transition in the cesium atom. The first clocks achieved an uncertainty of about 10^{-10}. By steady research and refinement, the clocks were improved until their uncertainty reached the level of about 5×10^{-16}. However, it is generally agreed that major improvements in cesium clocks are no longer likely. Fortunately, a new type of clock is now being realized.

The timing element in an atomic clock is the frequency of a transition between energy levels in an atom or ion. The measured precision of the clock is proportional to the size of the transition frequency, assuming that the ability to measure the frequency is kept the same. Because optical frequencies are larger than microwave frequencies by a factor of ~10^5, optical clocks hold the potential of being enormously more precise than the cesium clock.

Rosenband *et al.* report the comparison of two atomic clocks based on the frequencies of optical transitions in single ions. One clock uses the Al^+ ion, the other uses the Hg^+ ion. They measured the ratio of the frequencies of the two clocks to an uncertainty of 5.2×10^{-17}. This result is among the most precise measurements ever made in physics.

The clocks used by Rosenband *et al.* employ a single ion that is confined in a trap by electric fields. The experimental challenge is to approach as closely as possible the ideal of a single particle at rest in space, free from all perturbations and measured as well as quantum mechanics permits. They use a series of imaginative techniques that have been developed by themselves and others. As with every high-precision measurement, the principal challenge was to evaluate the effects of perturbations and sources of uncertainty. Although the sources are quite different for the two clocks, their final uncertainties are approximately the same, yielding an overall uncertainty of 5.2×10^{-17}.

When precision is pushed to new levels, ever more subtle effects must be taken into account. For instance, the error budget includes a small contribution, 1×10^{-18}, due to an uncertainty in the gravitational potential of the two clocks. This corresponds to a difference in their altitudes of 1 cm. This heralds one of the most interesting aspects of time keeping with optical clocks: The effects



Comparing clocks. Lasers (red) are frequency-locked to atomic transitions in each laboratory, forming independent atomic clocks. Another laser (green) is used to communicate between the clocks over an optical fiber. The frequencies of the red and green beams are controlled and compared by using a femtosecond frequency comb in each laboratory.

CREDITS: (TOP) T. ROSENBAND/NATIONAL INSTITUTE OF STANDARDS AND TECHNOLOGY (BOTTOM) J. YE/JILA

Department of Physics, Massachusetts Institute of Technology (MIT), and MIT–Harvard Center for Ultracold Atoms, Cambridge, MA 02139, USA. E-mail: kleppner@mit.edu

of general relativity that mix time with gravity are starting to approach a point that will require rethinking the basic concept of "keeping time."

Clockmakers face a second dilemma: The more accurate clocks are, the more difficult it is to compare them. The clocks used by Rosenband *et al.* were located in the same building, at the National Institute of Standards and Technology (NIST) in Boulder, Colorado, and compared with fiber links of a few hundred meters. Comparing clocks that are very far apart presents a separate challenge. Precision timing signals between distant laboratories are currently transmitted over microwave networks or by satellites, but these fail at the levels of precision now being achieved with optical clocks.

Ludlow *et al.* describe the comparison of clocks that employ optical transitions in neutral atoms, in contrast to ions. Furthermore, the clocks are separated by kilometer-scale urban distances (see the second figure). One clock was located at JILA on the campus of the University of Colorado in Boulder, and the other at NIST just outside the campus. The JILA clock employed strontium atoms held by light waves; the NIST clock used calcium atoms that were unconfined but so cold that their velocities were relatively low. These clocks also outperformed cesium clocks, and their rates were compared with a fractional uncertainty at the level of 10^{-16}. Most notably, by comparing clocks in different laboratories with optical signals transmitted by fiber in contrast to microwave communications, this work represents an important advance in time transfer at the frontiers of precision.

Ion-based atomic clocks currently achieve the highest accuracy because of their relative freedom from perturbations. However, neutral atom-based atomic clocks offer the advantage of much stronger signals, because the ion clocks use only a single particle whereas atom clocks typically use tens of thousands of atoms. There are numerous candidates for the new generation of optical atomic clocks, and eventually the second will be redefined based on one of them. However, that is unlikely to happen soon, because currently there is no obvious best choice for an ion or atom optical clock.

The advances in optical clocks described by Rosenband *et al.* and Ludlow *et al.* represent a milestone in time keeping because both groups achieved uncertainties that are significantly below those of primary cesium time standards. These realizations of optical atomic clocks rest on developments that stretch back more than 20 years. Enabling technologies include methods for trapping and cooling single ions developed by Wineland and his collaborators in the 1980s (*3*); laser cooling atoms for which Chu, Cohen-Tannoudji, and Phillips received the Nobel Prize in 1997 (*4*–*6*); the development of methods for ultrahigh optical and ultraviolet spectroscopy of ions by Bergquist and his collaborators in the 1990s (*3*); and the invention of the femtosecond frequency comb and optical frequency metrology for which Hänsch and Hall received the Nobel Prize in 2005 (*7*, *8*).

It will take some time to engineer an optical clock so that it can operate with the reliability and simplicity needed for practical applications, but once the goal is clearly in sight, this sort of engineering can move speedily. The question inevitably arises as to what the next generation of clocks will be useful for. One can point to basic tests such as the constancy of the fundamental constants, and possible applications such as geodesy. However, the best response to that question is simply to note that when atomic clocks were invented 50 years ago, nobody was dreaming of the Global Positioning System (GPS). The development of the GPS illustrates the truth of the adage that revolutionary technologies are likely to generate revolutionary applications.

References

1. T. Rosenband *et al.*, *Science* **319**, 1808 (2008); published online 6 March 2008 (10.1126/science.1154622)
2. A. D. Ludlow *et al.*, *Science* **319**, 1805 (2008); published online 14 February 2008 (10.1126/science.1153341)
3. J. C. Bergquist, S. R. Jefferts, D. J. Wineland, *Physics Today* **54** (3), 37 (March 2001).
4. S. Chu, *Rev. Mod. Phys.* **70**, 685 (1998).
5. C. Cohen-Tannoudji, *Rev. Mod. Phys.* **70**, 707 (1998).
6. W. D. Phillips, *Rev. Mod. Phys.* **70**, 721 (1998).
7. T. W. Hänsch, *Rev. Mod. Phys.* **78**, 1297 (2006).
8. J. L. Hall, *Rev. Mod. Phys.* **78**, 1279 (2006).

10.1126/science.1155948

ECONOMICS

When a Commodity Is Not Exactly a Commodity

Nancy Folbre

Economic transactions for services such as health and elder care are complicated by personal interactions and emotional connections.

The metabolism of our global economy relies on trillions of daily transactions, many of which involve goods and services termed commodities. In the idealized competitive markets of conventional economic theory, specific commodities are homogeneous and their quality is easily assessed. Because all pork bellies are alike, your choices among them can be based on price alone. But market transactions are not always so straightforward. Researchers have recently begun to explore the ways in which the process of exchange itself may modify the exchangers—altering product quality in unanticipated ways.

Labor, for example, is not a perfect commodity because its quality can be profoundly affected by the motivation of the laborer. The difficulty of controlling motivation makes labor contracts incomplete—the services being purchased cannot be perfectly specified in advance and are subject to change.

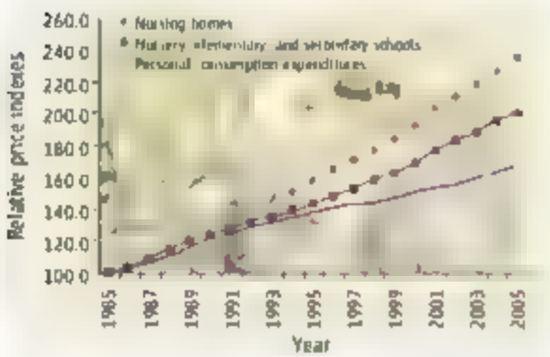
John Maynard Keynes noted that employers whose demand for labor falls often prefer layoffs to nominal wage reductions, which can elicit retaliation in the form of reduced effort (*1*). Mathematical models of the strategic relationship between individuals (principals) trying to control the behavior of others (agents) show that employers may pay a wage premium or offer gifts to workers in order to elicit greater effort (*2*–*4*). And a growing empirical literature based on laboratory experiments shows that workers' perceptions of fairness may also affect effort (*5*). In other words, wage-cutting can backfire.

These complexities of employer-worker interaction can be compounded by interactions between workers and consumers. Such interactions are relatively uncommon in the purchase of agricultural products or manufactured goods, where sales are typically distant from the point of production. The purchase of services, however, often entails high levels of personal interaction. The emotional dimension of work is particularly salient in services that entail provision of care for dependents—children, the elderly, and the sick or disabled.

Health, education, child care, and elder care account for a growing percentage of paid employment in countries like the United

Department of Economics, University of Massachusetts, Amherst, MA 01003, USA. E-mail: folbre@econs.umass.edu

States, particularly among women (*6*). Care services also account for an increasing share of personal consumption expenditures. Prices in the "care sector" have been rising faster than in other parts of the economy, and concerns about the quality of care services have also been escalating (*7*). Many factors, including rapid technological change, account for these trends. But even in two labor-intensive industries away from the leading edge of technical innovation—nursing homes and nursery elementary secondary schools · prices have increased more rapidly than overall personal consumption expenditures (see the figure)



Costs of care. Price indices for personal consumption (solid line), education (squares), and nursing home care (circles), relative to the year 1985. Data from (*17*).

The distinctive characteristics of care work itself help explain this phenomenon.

Almost by definition, care is difficult to standardize and therefore resists commodification. Most of us know the names of the doctors and nurses caring for us and hope they will remember ours, at least until we get well. Children learn their teachers' names, and vice versa. Unlike many other services, successful care work often requires collaboration between worker and consumer. Doctors often must cajole patients to take medication or change their life-styles, teachers must persuade students to study. Child care and elder care workers require a high level of personal cooperation from their wards. Care services have idiosyncratic and emotionally sticky aspects that affect both their cost and their quality—people's preferences and feelings can be modified by the very process of exchange (*7*). They form attachments.

Employers, insurance companies, and other institutions that mediate relationships between care workers and consumers often face pressure to cut care costs. From the vantage point of these intermediaries, the "stickiness" of care work makes it difficult to increase efficiency. Doctors who are emotionally involved with their patients may overprescribe expensive treatments, teachers who love their students may be unwilling to impose harsh discipline. This concern provides a rationale for impersonal administrative rules that remove decision-making authority from direct care providers and for emphasis on standardized quality measures such as length of hospital stay or performance on standardized tests.

But "stickiness" also has positive effects, precisely because care doesn't fit the characteristics of a standard commodity. Sustained personal interaction with care recipients can strengthen intrinsic motivation to help them, which enhances performance in complex jobs characterized by task ambiguity, where it is difficult to specify in advance which specific goals should take priority (*8*). Further, emotional connection often yields valuable person-specific information. Unlike the idealized consumers of standard economic theory, care recipients may not know themselves what they need, and they don't enjoy a menu of many alternative choices. Their very lack of consumer sovereignty makes them vulnerable to institutional pressures to cut costs by lowering aspects of quality that are difficult to measure.

An important class of models describing the strategic relationship between principals and agents (such as employers and their employees) shows that pressure to reallocate effort to easily measured aspects of performance is likely to result in a reallocation of effort away from less easily measured aspects of performance (*9*). Evidence suggests that those who enter caring occupations do so in part because they derive personal satisfaction from meeting the needs of other individuals (*10*). Working conditions or performance standards that make it difficult for them to do so can encourage exit from—or discourage entry into—caring professions.

Changes in gender roles are contributing to increases in the cost of care reflected in the figure. In the past, restrictions on women's access to education and employment channeled many into relatively low-paying jobs in the care sector. Their rapid movement into high-level professional and managerial jobs in recent years has shifted their labor supply and also increased the demand for services such as child care and elder care. Women also have increased bargaining power as managers and consumers, where they tend to be particularly attentive to concerns about quality of care because they are fall-back providers in the family.

Baumol famously argued years ago that most services were doomed to slow productivity growth and rising costs relative to manufacturing sectors (*11*). The growth of information technology has proved him partly wrong—labor productivity in many services, including finance, insurance, and retail sales, has soared. New information technology can also improve the productivity of many care services through innovations such as standardized medical records and online courses. Research on "carebots" designed to automate some aspects of elder care is well under way (*12*).

But our global economic metabolism depends on hearts as well as minds. Most care services will retain an important personal and emotional dimension, even when they are bought and sold. Further, the long-run efficiency of purchased care services will be shaped by their articulation with unpaid family-provided services whose value is not factored into standard national income accounts.

Do small-scale nursing homes do a better job ensuring care quality than larger ones (*13*)? Can institutions increase service quality by reducing labor turnover and encouraging consistent long-term assignment of caregivers to particular care recipients (*14*)? Does family instability contribute to poor health and education outcomes for children (*15*)? What kinds of internet support groups best help individuals cope with health problems (*16*)? More research on the impact of personal interaction and emotional connection in provision of care services could contribute to improved institutional design.

References

1. J. M. Keynes, *The General Theory of Employment, Interest, and Money* (Macmillan, London, 1936).
2. G. Akerlof, *Q. J. Econ.* **97**, 4 (1982).
3. S. Bowles, *Am. Econ. Rev.* **75**, 1 (1985).
4. J. E. Stiglitz, *J. Econ. Lit.* **25** (1987).
5. S. Gächter, E. Fehr, in *Surveys in Experimental Economics. Bargaining, Cooperation and Election Stock Markets*, F. Bolle, M. Lehmann-Waffenschmidt, Eds. (Physica-Verlag, Heidelberg, 2001), pp. 95–132.
6. N. Folbre, J. Nelson, *J. Econ. Perspect.* **14**, 4 (2000).
7. N. Folbre, *Polit. Soc.* **34**, 1 (2006).
8. D. M. Kreps, *Am. Econ. Rev.* **87**, 2 (1997).
9. B. Holmstrom, P. Milgrom, *Am. Econ. Rev.* **84**, 4 (1994).
10. P. England, M. Budig, N. Folbre, *Soc. Probl.* **49**, 4 [illegible].
11. W. J. Baumol, *Am. Econ. Rev.* **57**, 3 (1967).
12. J. Pineau, M. Montemerlo, M. Pollack, N. Roy, S. Thun, *Robot. Auton. Syst.* **42**, 271 (2003).
13. R. Kane, T. Y. Lum, L. J. Cutler, H. W. Degenholtz, T. Yu, *J. Am. Ger. Soc.* **55**, 6 (2007).
14. N. G. Castle, J. Engberg, A. Men, *The Gerontologist* **47**, 650 (2007).
15. P. Fomby, A. Cherlin, *Am. Sociol. Rev.* **72**, 181 (2007).
16. A. Barak, *Comput. Hum. Behav.* **23**, 971 (2007).
17. Statistical Abstract of the United States, 2008, Table 718. Price Indexes for Personal Consumption Expenditures by Type of Expenditure: 1929 to 2005, available online at www.census.gov/compendia/statab/cats/prices.html.

10.1126/science.1153904

ATMOSPHERIC SCIENCE

Recording Earth's Vital Signs

Fifty years ago, continuous measurements of atmospheric carbon dioxide were begun at Mauna Loa, Hawai

Ralph F Keeling

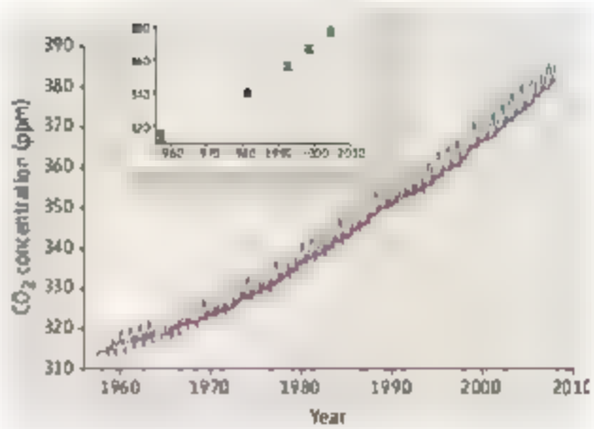
This year marks the 50th anniversary of the start of the Mauna Loa CO_2 record, the longest continuous record of CO_2 in the atmosphere. Initiated by my father, Charles D. Keeling of the Scripps Institution of Oceanography, the record provided the first compelling evidence that the concentration of CO_2 in the atmosphere was rising. It has become an icon of the human imprint on the planet and a continuing resource for the study of the changing global carbon cycle. The Mauna Loa story (*1*) provides a valuable lesson on the importance of continuous Earth observations in a time of accelerating global change.

At the outset, the decision to place the instrument at Mauna Loa was a gamble. Existing measurements suggested that atmospheric CO_2 concentrations varied widely depending on the place and time. Given this variability, could a meaningful record be recovered from an instrument parked in one location? Among the skeptics was Roger Revelle, then director of the Scripps Institution of Oceanography. Revelle would eventually become one of the record's strongest champions. Initially, however, he urged that priority be given to a one-time survey of CO_2 variability using ships and airplanes. Such a survey could be repeated a decade or so later to look for long-term changes.

My father was armed with evidence from his postdoctoral research that the CO_2 concentration in the remote atmosphere was a lot less variable than previously believed (*2*). He also had a strong ally in Harry Wexler of the U.S. Weather Bureau, who envisioned a central role for the newly established Mauna Loa Observatory (see the photo) in the major field program planned for the International Geophysical Year of 1957–1958.

The value of the Mauna Loa data soon was apparent (*3*). By the second year, a regular seasonal cycle was evident, reflecting the "breathing" of land plants in the Northern Hemisphere. Together with a more limited CO_2 data set from the South Pole, begun in 1957, the record documented a global rising trend attributable to the burning of fossil fuels worldwide (see the graph). In the 1960s and 1970s, the curve was seen by countless scientists, some of whom were drawn to study the science of global warming by the curve's ominous rise.

What if CO_2 had been measured only via repeated global surveys, as envisioned by Revelle? As the inset in the graph shows, a CO_2 record degraded to include only one point every decade or two loses its convincing message. Variations from survey to survey may be instrumental artifacts, or the apparent trend may be a random fluctuation. As a recent study of ocean currents in the North Atlantic has shown, resolving trends from repeated surveys can be perilously difficult (*4*).



The value of continuous data. Within a few years, the continuous Mauna Loa (blue) and South Pole (red) records provided convincing evidence that CO_2 was rising. If CO_2 had been measured only as often as surveys of the North Atlantic overturning circulation (*4*), it would have taken decades to obtain convincing evidence (inset).

The Scripps CO_2 program was shut down briefly in 1964 following congressionally mandated budget cuts. A more serious challenge loomed in the 1970s, when my father was asked to draw a line between the part of the CO_2 program that was basic research and the part that constituted "routine monitoring"; the routine activities would be transferred to a government agency. My father did not comply with the request.

In the 1970s, the Scripps CO_2 program expanded to an array of eight stations distributed globally. A large part of the effort was being expended not in routine data collection, but in the messier process of identifying and eliminating systematic errors. As the records grew, additional features emerged, such as a link between interannual CO_2 fluctuations and El Niño events (*5*) and changes in the amplitude of the seasonal cycle with time (*6*). Were these features real or artifacts? There was no way to be sure without revisiting the fundamentals of the instrumentation, performing new calibrations, and reprocessing all of the records. This cycle was repeated many times as scientific interests evolved.

The distinction between research and routine monitoring may seem clear when applied to an activity like weather forecasting, but in the case of a program aimed at tracking long-term change, "research" and "operations" cannot be separated cleanly. Finding and correcting for the inevitable systematic biases is a job for scientists who understand the measurement technology, are passionate about data integrity, and are motivated to unravel how the Earth system operates.

The Mauna Loa experience also illustrates the critical need for redundancy. From the outset the Mauna Loa record was backed by the parallel record from the South Pole. In 1960, a second record was begun at Mauna Loa, based on flasks shipped back to Scripps for analysis. It is an inescapable fact that if you are trying to track changes over time, you only get one chance to measure each point. To prove you got it right, you must take measurements in multiple ways. And the challenge may come decades later. A recently

The Mauna Loa Observatory.

PHOTO CREDIT: FORREST MIMS III/WWW.FORRESTMIMS.ORG

Scripps Institution of Oceanography, University of California, San Diego, La Jolla, CA 92093, USA. E-mail: rkeeling@ucsd.edu

discovered slowing of the rate of CO_2 uptake by the Southern Ocean appears to hinge on questionable CO_2 measurements made at Ascension Island in the 1980s (*7*, *8*). A duplicate record would have settled the issue.

The Scripps CO_2 program is now a component of a multinational collaboration aimed at tracking changes in greenhouse gases and related species, coordinated by the World Meteorological Organization. A major justification of this effort is the promise of quantifying the sources and sinks of greenhouse gases at Earth's surface. Soon, the compliance of international treaties to curb greenhouse gas emissions may be assessed using these capabilities.

If long-term observations are fundamental to understanding global change, why have they proved so hard to support? The costs of sustained measurements can be high, so prioritization is clearly an issue. The Scripps program has proved, however, that a long-term observational program is not necessarily incompatible with the normal peer review system. The Scripps program continues to be funded—if perilously—one grant at a time. Even within agencies committed to long-term observations, such as the National Oceanographic and Atmospheric Administration, funding is tight and a hiatus may be only one political wind shift or economic downturn away. A diversity of funding sources supporting a heterogeneous mixture of overlapping programs is probably the best formula for long-term stability.

A continuing challenge to long-term Earth observations is the prejudice against science that is not directly aimed at hypothesis testing. At a time when the planet is being propelled by human action into another climate regime with incalculable social and environmental costs, we cannot afford such a rigid view of the scientific enterprise. The only way to figure out what is happening to our planet is to measure it, and this means tracking changes decade after decade and poring over the records. A point of diminishing scientific returns has never been realized in what is now known as the "Keeling Curve," the Mauna Loa CO_2 record.

References

1. C. D. Keeling, *Annu. Rev. Energy Environ.* **23**, 25 (1998).
2. C. D. Keeling, in *Proceedings of the Conference on Recent Research in Climatology*, W. Craig, Ed. (Committee on Research in Water Resources and University of California, Scripps Institution of Oceanography, La Jolla, CA, 1957), pp. 43–49.
3. C. D. Keeling, *Tellus* **12**, 200 (1960).
4. H. L. Bryden, H. R. Longworth, S. A. Cunningham, *Nature* **438**, 655 (2005).
5. R. B. Bacastow, *Nature* **261**, 116 (1976).
6. R. B. Bacastow, C. D. Keeling, T. P. Whorf, *J. Geophys. Res.* **90**, 10529 (1985).
7. R. M. Law, R. J. Matear, R. J. Francey, *Science* **319**, 570a (2008).
8. C. Le Quéré *et al.*, *Science* **319**, 570c (2008).

10.1126/science.1156761

BIOCHEMISTRY

A Postgenomic Visual Icon

A decade of experience in visualizing large-scale genotypic and phenotypic data as heat maps has illuminated the strengths and limitations of the approach.

John N. Weinstein

The "postgenomic era" in biology may be hard to define, and when it actually began is subject to debate. But its most characteristic feature is clearly the accumulation of massive amounts of genotypic and phenotypic data that must be organized, analyzed, visualized, and interpreted. That series of challenges has been central to recent bioinformatics. For visualization, by far the most popular graphical representation has been the "clustered heat map," which compacts large amounts of information into a small space to bring out coherent patterns in the data. Despite its popularity, however, are such maps optimal for visually integrating information to extract valuable insights and generate fresh hypotheses? That question can be addressed through understanding the strengths and limitations of heat map visualization.

Since their debut over 10 years ago (*1*) (see the figure), clustered heat maps have appeared in well over 4000 biological or biomedical publications. They have been used for two-dimensional display of patterns in all types of molecular data, including messenger RNA (mRNA) and microRNA expression, protein expression, DNA copy number, DNA methylation, metabolite concentration, and drug activity (*1*–*8*). They have proved useful for microarray data (*2*) and have sometimes been engineered for "integromic" merging (*1*, *9*, *10*) of different types of molecular information. The figure, for example, combines data on mRNA expression, protein expression, mutations, cell cycle properties, stress responses, a yeast-based functional assay, and drug activity in cancer cells. Organisms analyzed have spanned the phylogenetic tree from the plant *Arabidopsis thaliana* to rainbow trout to suicidal crickets (*11*–*13*). Diseases analyzed have ranged from AIDS to cancer to bubonic plague (*1*, *14*, *15*).

In the case of gene expression data, the color assigned to a point in the heat map grid indicates how much of a particular RNA or protein is expressed in a given sample. The gene expression level is generally indicated by red for high expression and either green or blue for low expression. Coherent patterns (patches) of color are generated by hierarchical clustering on both horizontal and vertical axes to bring like together with like. Cluster relationships are indicated by tree-like structures adjacent to the heat map, and the patches of color may indicate functional relationships among genes and samples. Occasionally, a source of order other than clustering (for example, time in a series of measurements) is used on one or both axes. Without some basis for functional ordering on both axes, however, there would be no coherent patterns of color.

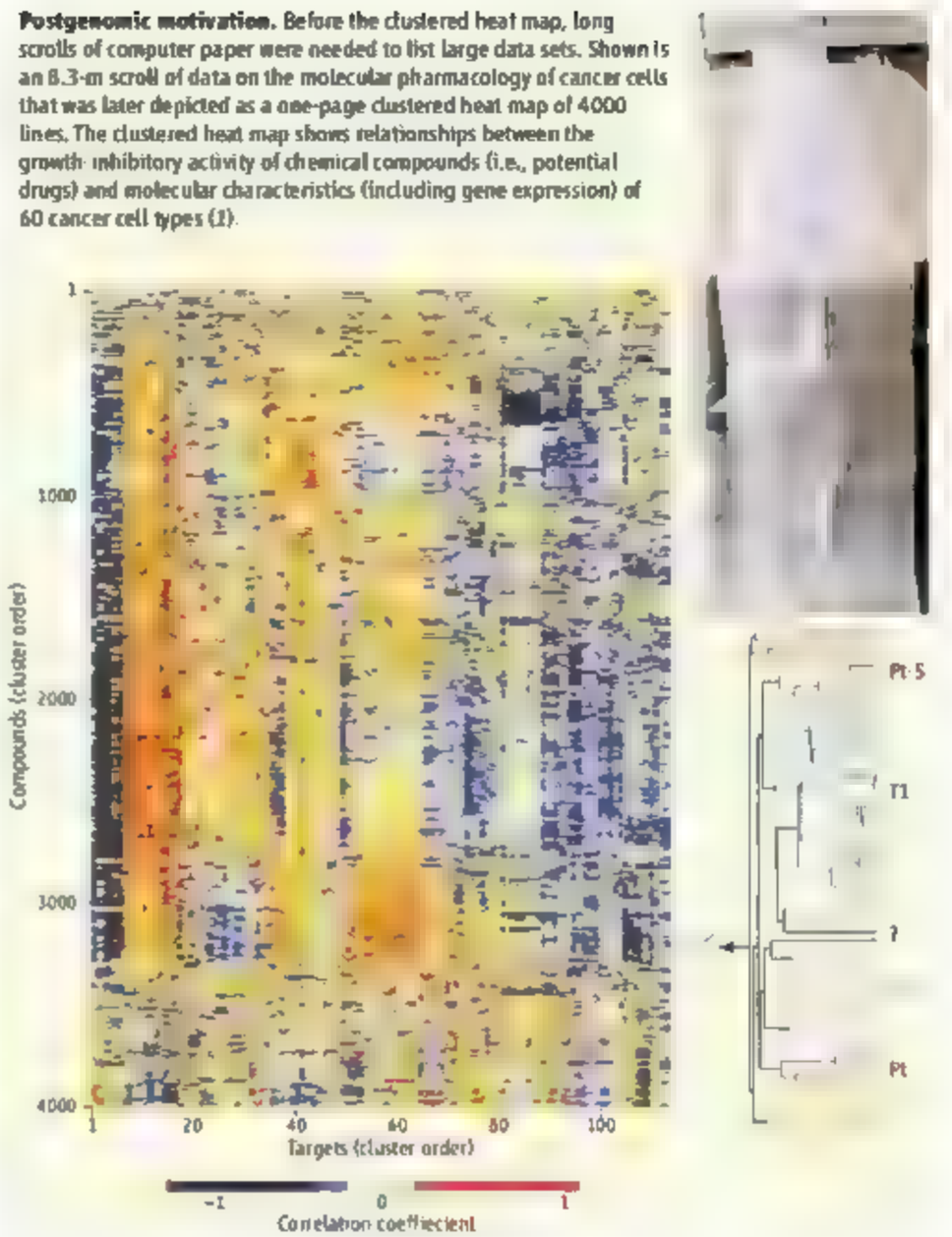
Seductive though it may be, the clustered heat map has its limitations and potential for misinterpretation or misuse. Most prominently among the limitations, it provides only first-order insight into the data; complex patterns of nonlinear relationship among only a few of the samples are unlikely to show up. A computer-intensive variant based on "biclustering" has been developed to reveal such relationships (*16*). A second problem is that, in hierarchical clustering, each bifurcation of the cluster tree can be "swung" in either direction at each fork in the tree, so some objective (but, to a degree, arbitrary) rule must be invoked to decide which way each branch will, in fact, swing. There is also the temptation to select a small subset of the variables (for example, genes in a microarray study), and represent them in a clustered heat map. That is common (and appropriate) practice in the discovery of new biomarkers and gene expression signatures for discriminating subtypes of a disease such as cancer (*17*). However, if one picks a signature consisting of only a few dozen genes out of a set of more than 10,000, then even randomized

Department of Bioinformatics and Computational Biology, M. D. Anderson Cancer Center, 1515 Holcombe Boulevard, Houston, TX 77030, USA. E-mail: jweinste@mdanderson.org

data can produce clustered heat maps that appear spuriously to show good distinction of two subclasses. Typically, in such small-subset "cherry-picking" analysis, the upper left and lower right quadrants of the figure will be predominantly one color (for example, red), whereas the lower left and upper right quadrants will be the other color (for example, green or blue). It may be legitimate to show such maps for gene signatures simply because the entire set of genes cannot be annotated on a single map. But the lack of statistically robust meaning to the pattern should be clearly stated, and that is rarely done.

Even beyond those limitations and concerns, the generation of clustered heat maps is a surprisingly subtle process. It requires that a large number of choices be made, and those choices dictate the type and meaning of pattern that emerges. The required decisions include (i) the preprocessing algorithm (e.g., type of background subtraction, data normalization, and data filtering), which hopefully minimizes noise in the system while keeping the meaningful signal, (ii) the clustering algorithm (e.g., average linkage, complete linkage, or centroid-based), which determines how the data will be grouped; (iii) the distance metric (e.g., Euclidean or correlation), which defines what is meant by similarity of genes or samples to each other; and (iv) the color scheme (linear, logarithmic, quantile, two-color, or three-color), which determines what patterns in the data will be emphasized visually. The decision to use relative or absolute data must also be made, and one data set can be subtracted from another to create a "difference" heat map (e.g., of gene expression before and after treatment of cells with a drug). [See supporting online material and (*1*, *18*) for more detailed explanations of those choices and their implications.] The point is that many different heat maps, each with its own visual meaning, can be generated from the same experiment. Unless the details of the parameter choices are specified, the analysis is incomplete at best and subject to misinterpretation at worst.

Postgenomic data sets are becoming larger and larger. A decade ago, microarrays produced thousands of numbers; now they often produce millions. Hardware, software, mathematical algorithms, and the human mind are all being stressed to the limit by the flood of data. Every biomedical research institution is finding its human and computational resources inadequate to the task. Our need for graphical representations that illuminate patterns in the data and evoke new hypotheses will only increase over the next few years, particularly in the drive to identify biomarkers useful for "personalizing" treatment of diseases such as cancer. The ability of the human eye to recognize patterns, coupled with modern data analysis, "can stimulate excitement, awe, new ways of looking at things, and, above all, a broad appreciation of even the most esoteric scientific information" (*19*). For the last decade, the ubiquitous clustered heat map has served those purposes, even if imperfectly, for postgenomic biology.



Postgenomic motivation. Before the clustered heat map, long scrolls of computer paper were needed to list large data sets. Shown is an 8.3-m scroll of data on the molecular pharmacology of cancer cells that was later depicted as a one-page clustered heat map of 4000 lines. The clustered heat map shows relationships between the growth-inhibitory activity of chemical compounds (i.e., potential drugs) and molecular characteristics (including gene expression) of 60 cancer cell types (*1*).

References

1. J. N. Weinstein *et al.*, *Science* **275**, 343 (1997).
2. M. B. Eisen, P. T. Spellman, P. O. Brown, D. Botstein, *Proc. Natl. Acad. Sci. U.S.A.* **95**, 14863 (1998).
3. D. T. Ross *et al.*, *Nat. Genet.* **24**, 227 (2000).
4. U. Scherf *et al.*, *Nat. Genet.* **24**, 236 (2000).
5. G. Szakács *et al.*, *Cancer Cell* **6**, 129 (2004).
6. B. R. Zeeberg *et al.*, *BMC Bioinformatics* **6**, 168 (2005).
7. S. Nishizuka *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **100**, 14229 (2003).
8. M. J. Brauer *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **103**, 19302 (2006).
9. T. G. Myers *et al.*, *Electrophoresis* **18**, 467 (1997).
10. J. N. Weinstein, Y. Pommier, *C. R. Biol.* **326**, 909 (2003).
11. J. Bobe, J. Montfort, T. Nguyen, A. Fostier, *Reprod. Biol. Endocrinol.* **4**, 39 (2006).
12. A. C. Oosey *et al.*, *Mol. Biol. Evol.* **24**, 1045 (2007).
13. O. G. Biron *et al.*, *Insect Mol. Biol.* **15**, 731 (2006).
14. Y. Le Proust *et al.*, *J. Immunol.* **177**, 5145 (2006).
15. X. Y. Wang *et al.*, *Arch. Microbiol.* **186**, 151 (2006).
16. Y. Kluger, R. Basri, J. T. Chang, M. Gerstein, *Genome Res.* **13**, 703 (2003).
17. T. R. Golub *et al.*, *Science* **286**, 531 (1999).
18. http://discover.nci.nih.gov/cimminer
19. J. Nesbit, M. Bradford, *Science* **317**, 1857 (2007).

Supporting Online Material
www.sciencemag.org/cgi/content/full/319/5871/1772/DC1

10.1126/science.1151888

BOSTON 2008

Annual Meeting Explores Cooperation, Competition in Global Science Efforts

BOSTON—AIDS is still the number one cause of death in Africa and seventh largest cause of death worldwide, but recent success in fighting the epidemic is a hopeful sign that cooperation between scientists and policy-makers can address problems like HIV/AIDS on a global scale, according to experts at the 2008 AAAS Annual Meeting.

Competitively priced antiretroviral drugs from new production centers in places like India, along with the work of international collaborations to distribute those drugs, have expanded HIV treatment significantly in the developing world, said AAAS Chair David Baltimore, the panel's moderator. Today, more than 3 million HIV/AIDS patients in the developing world receive antiretroviral medications, compared to 200,000 patients in 2001.

The complex mix of national competition and international cooperation that defines the response to the HIV/AIDS crisis is indicative of how science is practiced around the world. Baltimore suggested in his presidential address to open the Boston meeting. Under the banner "Science and Technology from a Global Perspective," the meeting brought leading scientists, engineers, educators, and policymakers from 56 countries together to discuss how the twin themes affect cutting-edge research in ocean pollution, nuclear smuggling, climate change, and science education.

International cooperation has been a tremendous benefit to infectious disease researchers, particularly those working on HIV/AIDS. Plenary speaker Jim Yong Kim, director of the Francois Xavier Bagnoud Center for Health and Human Rights at the Harvard School of Public Health, called the response to the AIDS epidemic "the first time in history that the wealthiest, most powerful people have committed to chronic care for chronic conditions to the poorest people on earth."

But Kim and the other speakers in the global health panel said scientists need to do more to bring the best HIV care into communities. And at a breakfast for journalists on the first day of the meeting, Baltimore said scientists were "not any closer than we were in the 1980s" to developing an HIV vaccine—a startling comment to many of the reporters. The insight will "inform my future coverage of AIDS," said Clive Cookson, science editor of the *Financial Times*.

(L-R) Jim Yong Kim and David Baltimore

CREDIT: KOJELLA/PHOTO.COM

In several sessions, participants discussed whether international cooperation might undermine individual countries' efforts to train and retain members of the S&T workforce. In a topical lecture on American-European science cooperation, the speakers acknowledged fierce competition between nations to lure scientists from the developing world and research-based businesses such as pharmaceutical companies. But global science "need not be thought of as an arms race," since countries tend to benefit as much from multinational research projects as they do from national projects, said Shirley Ann Jackson, president of Rensselaer Polytechnic Institute and a past president of AAAS.

Rwanda is one developing country that hopes to strengthen its economy through S&T investment and international cooperation. Rwandan President Paul Kagame, who shared the meeting's opening plenary with Baltimore, said his government plans to spend 5% of its gross domestic product on science and technology capacity by 2012. Repeating a request made last fall when Baltimore visited Rwanda, Kagame asked for help from AAAS's network of scientists in rebuilding his country's research capacity.

"We're absolutely committed to working together with [Kagame]," said AAAS CEO Alan I. Leshner in an interview with Robert Frederick, the host of *Science* Podcast. "Anybody who shows that level of dedication and commitment to the application of science for the advancement of a very poor country deserves our collaboration."

Boston's mix of high-profile scientists and policy leaders like Kagame impressed journalists like *Ars Technica* Science Editor John Timmer. "It was like a Davos meeting filled with people who tell the actual Davos attendees what's important," he wrote in his blog.

Indeed, a number of those who attended the Boston meeting were veterans of the 2008 Davos World Economic Forum, where the issue of climate change was discussed prominently. In their AAAS plenary speeches, Davos participants Nina Fedoroff, science and technology adviser to the U.S. Secretary of State, and Judith Rodin, president of the Rockefeller Foundation, urged scientists to become more involved in solutions that address the unequal burden of climate change around the world.

–Becky Ham

INTERNATIONAL

AAAS Strengthens S&T Ties with Vietnam

A new agreement between AAAS and Vietnam's National Institute for Science and Technology Policy and Strategy Studies strengthens AAAS's ties with the science and technology leaders of one of Asia's most dynamic economies while supporting further innovation in Vietnam through the promotion of science and scientists.

The memorandum of understanding, signed at AAAS Headquarters on 27 February, follows more than 2 years of high-level conferences and visits between AAAS officials and their Vietnamese counterparts. It formalizes the Association's plans to "collaborate in the advancement of science, technology, and innovation policy and studies," and to pursue joint activities.

AAAS Chief International Officer Vaughan Turekian signed the agreement along with Tran Ngoc Ca, the National Institute's vice director. Turekian said the presence of more than 30 Vietnamese officials at the signing was "a real indication of the depth and broadness of the relationship" between the United States and Vietnam on S&T issues.

Tran Quoc Thang, Vietnam's vice minister for science and technology, said the cooperative agreement will help "to contribute to our friendship between two nations."

One of the first projects under the new agreement will be a symposium planned by AAAS under the auspices of the Asia-Pacific Economic Cooperation (APEC) forum, to be hosted by Vietnam and co-sponsored by the governments of China, New Zealand, and the United States this fall. The symposium will focus on R&D priorities and developing public-private partnerships among APEC's 21 members.

–Ginger Pinholster and Earl Lane

Doped Nanocrystals

David J. Norris,[1]* Alexander L. Efros,[2] Steven C. Erwin[2]

The critical role that dopants play in semiconductor devices has stimulated research on the properties and the potential applications of semiconductor nanocrystals, or colloidal quantum dots, doped with intentional impurities. We review advances in the chemical synthesis of doped nanocrystals, in the theoretical understanding of the fundamental mechanisms that control doping, and in the creation of highly conducting nanocrystalline films. Because impurities can be used to alter the properties of nanoscale materials in desirable and controllable ways, doped nanocrystals can address key problems in applications from solar cells to bioimaging.

The use of intentional impurities, or dopants, to control the behavior of materials lies at the heart of many technologies. Doping is critical for semiconductors, which would otherwise be electrically insulating. For this reason, researchers have begun to explore how dopants can influence semiconductor nanocrystals, crystallites a few nanometers in size with unusual and size-specific optical and electronic behavior (*1*). Undoped nanocrystals can now be prepared from most common group IV, III-V, and II-VI semiconductors, as well as in different shapes such as nanorods, tetrapods, and nanowires (*2*). They can be easily manipulated as colloidal dispersions and used as building blocks to create a family of complex artificial solids (*3*).

Why Dope Semiconductor Nanocrystals?

The promise of nanocrystals as a technological material, for applications including wavelength-tunable lasers (*4*), bioimaging (*5*), and solar cells (*6*), may ultimately depend on tailoring their behavior through doping. Impurities can strongly modify electronic, optical, and magnetic properties of bulk semiconductors. A substitutional impurity with one more valence electron than the host atom it replaces can be ionized by thermal energy and donate its extra electron to the semiconductor (*n*-type doping). Similarly, an impurity with one less valence electron can provide an extra hole (*p*-type doping). These electrons or holes are then available as carriers of electrical current. For nanocrystals, where applications often require thin conducting films, the ability to introduce these carriers is essential.

Dopants can also strongly influence optical behavior. Although undoped nanocrystals are highly fluorescent with a color that depends on size, lasers based on this emission are intrinsically inefficient. Several approaches can improve this situation (*4*), and one possibility is to incorporate dopants that provide carriers. For example, the lasing threshold in CdSe nanocrystals can be reduced threefold by adding extra electrons (*7*). For bioimaging applications, fluorescent dopants may mitigate toxicity problems by producing visible or infrared emission in nanocrystals made from less-harmful elements than those currently used (*8*). During prolonged illumination, as in solar cells, doping may also protect against photooxidation. The energy from absorbed photons can be efficiently transferred to the impurity, quickly localizing the excitation and suppressing undesirable reactions on the nanocrystal surface (*8*).

For magnetic dopants, confinement of the impurity within the nanocrystal can enhance its interactions with other carriers or quantum mechanical spins [see review (*9*)]. Interesting consequences include individual doped nanocrystals that behave as spin filters (*10*). Carriers whose spins are aligned one way are conducted preferentially, while others are blocked, an effect that could be used in future spintronic devices.

New Issues at the Nanoscale

Dopants in nanocrystals lead to phenomena not found in the bulk because their electronic states are confined to a small volume. For example, *n*- or *p*-type dopants can auto-ionize without thermal activation. This occurs because a carrier inside the crystallite must occupy one of the confined electronic states, which increase in energy with decreasing nanocrystal size (*1*). Below a

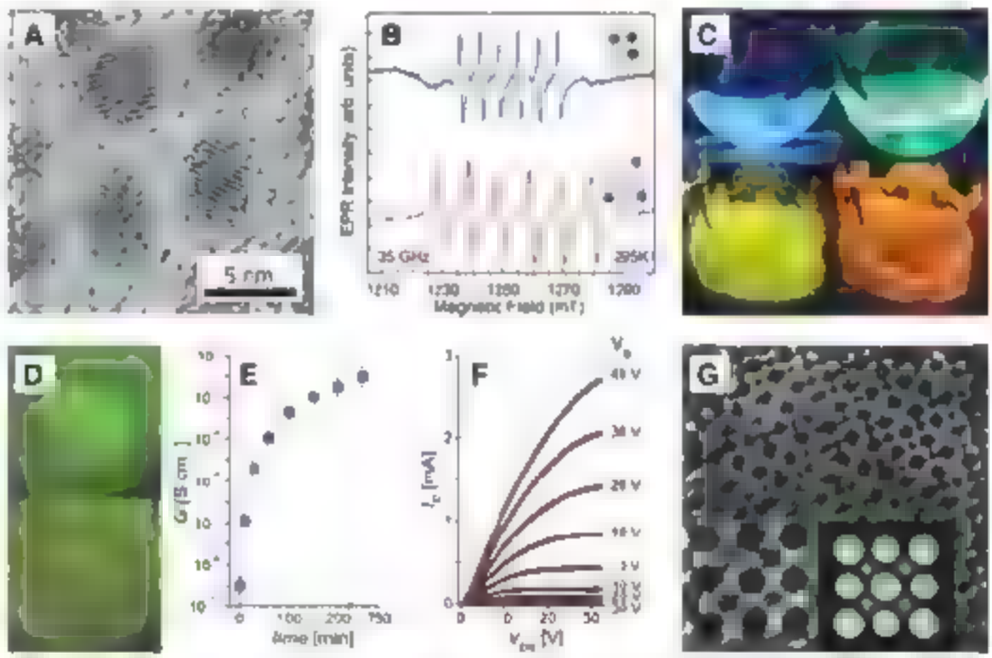


Fig. 1. (**A**) Transmission electron micrograph of 5-nm Mn-doped ZnSe nanocrystals. (**B**) Electron paramagnetic resonance spectra for ZnSe nanocrystals with Mn inside the crystallites (blue) and bound to their surfaces (red). The Mn resonance is split into six lines by the hyperfine interaction, which is sensitive to the local environment of the impurity. [Reprinted with permission from (*18*). Copyright 2001, American Chemical Society.] (**C**) Fluorescence of Cu- (top two images) and Mn- (bottom two images) doped ZnSe nanocrystals. The position of the impurity inside the nanocrystal changes the fluorescence color. [Reprinted in part with permission from (*8*). Copyright 2005, American Chemical Society, and courtesy of X. G. Peng, University of Arkansas.] (**D**) Fluorescence from undoped CdSe nanocrystals in an electrochemical cell (top). Electrons injected with an applied voltage (bottom) quench the emission. [Reprinted with permission from (*34*).] (**E**) Conductance of a film of undoped PbSe nanocrystals during exposure to hydrazine, an electron donor. [Reprinted with permission from (*36*).] (**F**) Current-voltage characteristics of a field-effect transistor made from a film of 8-nm PbSe nanocrystals. The drain current I_D versus the source drain voltage V_{DS} can be modulated by a gate voltage, V_G. [Reprinted with permission from (*36*).] (**G**) Transmission electron micrograph of a crystal (or superlattice) of two types of nanocrystals. (Inset) The arrangement of the 6.5-nm PbTe (purple) and 10.1-nm Ag_2Te (green) crystallites. [Reprinted by permission from Macmillan Publishers Limited: *Nature Materials* (*40*), copyright 2007.]

[1]Department of Chemical Engineering and Materials Science, University of Minnesota, Minneapolis, MN 55455, USA. [2]Center for Computational Materials Science, Naval Research Laboratory, Washington, DC 20375, USA.

*To whom correspondence should be addressed. E-mail: dnorris@umn.edu

critical radius, the confinement energy exceeds the Coulomb interaction between the ionized impurity and the carrier (*11*), which then automatically occupies a nanocrystal state.

Although the conduction of such carriers in bulk semiconductors is limited mainly by scattering from defects, in nanocrystal films it is controlled by the communication between crystallites. For high conductivity, neighboring nanocrystals should be uniform in size to have resonant electronic states and be closely spaced to ensure their quantum-mechanical overlap.

Conductivity also requires high carrier concentrations. Unfortunately, a nanocrystal with an extra electron or hole can act as a strong reducing or oxidizing agent, respectively. Consequently, electrochemical reactions at the surface can consume the carrier. Indeed, the redox properties of nanocrystals were an early reason for their study (*12*, *13*). If such a reaction occurs, the carrier will not participate in conduction. For films of doped nanocrystals, avoiding these reactions is a challenge. One strategy to keep carriers inside nanocrystals would be to coat a crystallite core with a shell of another semiconductor having a wider band gap. However, this barrier would also electrically isolate the carriers and suppress conductivity.

Exploration of all of these issues depends on having reliable methods to incorporate impurities. A typical nanocrystal, 5 nm in diameter (Fig. 1A), consists of a few thousand atoms. Adding a single impurity implies a dopant concentration of the order of 10^{-3}, or 10^{19} cm^{-3}. In a bulk semiconductor, this would constitute extremely heavy doping. Nevertheless, the average number of impurities per crystallite is still small, and stochastic fluctuations in their location and number can be substantial (*9*). When the properties of individual nanocrystals are of interest, such fluctuations must be controlled and minimized. This requires a deep understanding of the doping process and accurate experimental tools to confirm that the desired materials have been obtained.

Synthesis and Characterization of Doped Nanocrystals

Nanocrystals are synthesized by combining molecular precursors that contain the constituent elements. Thermal decomposition of the precursors leads to nucleation and growth. This general approach has been used in solid-, gas-, and liquid-phase reactions (*3*, *14*). The most uniform crystallites have been obtained by using colloidal chemistry. In this case, surfactants in the solution are used to control growth and passivate dangling bonds. Although these colloidal techniques were first optimized for CdSe, they have been extended to many other materials.

The most common strategy for doping is to include a precursor containing the impurity in the synthesis. However, this does not always succeed, and the resulting nanocrystals must be carefully characterized to determine whether the impurity was incorporated. To date, only a few analytical techniques have shown the ability to distinguish impurities on the surface from those inside. For Mn, electron paramagnetic resonance reveals spin interactions that are sensitive to the local environment of the impurity (*15*). Figure 1B shows how the spectrum changes whether the Mn is incorporated or surface-bound. This difference was first used to verify Mn-doped ZnS (*16*). Since then, other Mn-doped nanocrystals, including CdS (*17*) and ZnSe (*18*, *19*), have also been confirmed. For Co impurities, optical absorption spectroscopy reveals electronic transitions that are sensitive to environment and has confirmed doping in CdS and ZnS nanocrystals (*9*, *20*). Magnetic circular dichroism is another powerful technique (*17*, *18*). It can reveal whether a magnetic dopant is incorporated by measuring the influence of its magnetic spin on the nanocrystal states.

In addition to the simple synthetic approach, doping has been attempted with use of more sophisticated strategies. For example, molecular fragments of semiconductors can be chemically synthesized that already contain the impurity. These clusters are then used as the sole precursor in the growth solution. Nanocrystals such as Co-doped CdSe have been reported (*21*).

Mn and Co provide excellent systems to study doping because they also have unique magnetic and optical properties (*9*). However, in II-VI nanocrystals they are isovalent with the atoms for which they substitute and so cannot provide extra electrons or holes. Attempts at heterovalent doping have begun only recently, including colloidal Mn-doped InAs (*22*) and Li-doped ZnO (*23*), as well as P- and B-doped Si grown in the solid phase (*24*).

Despite doping successes, puzzles have also arisen. Even when impurities are incorporated, their concentration is typically an order of magnitude less than in the growth solution (*19*). For Mn in II-VI semiconductors, which has high bulk solubility (tens of percent), the concentrations attained in nanocrystals are far smaller (1% or less). Indeed, early efforts to dope Mn into CdSe, the most-studied system, failed to incorporate any impurities (*25*). These unexplained results have led to theoretical efforts to understand the mechanisms that control doping.

Theoretical Models

Several doping mechanisms have been proposed, as summarized in Fig. 2. Turnbull argued that crystallites tend to be pure simply because they contain very few atoms (*26*). This argument is purely statistical and does not propose any new physics for small crystals (Fig. 2A). Moreover, it cannot explain the issues with Mn doping of II-VI nanocrystals. Mn is sufficiently soluble in the bulk that even very small nanocrystals should still contain many impurities.

An alternative explanation is that nanocrystals undergo "self-purification," that is, the

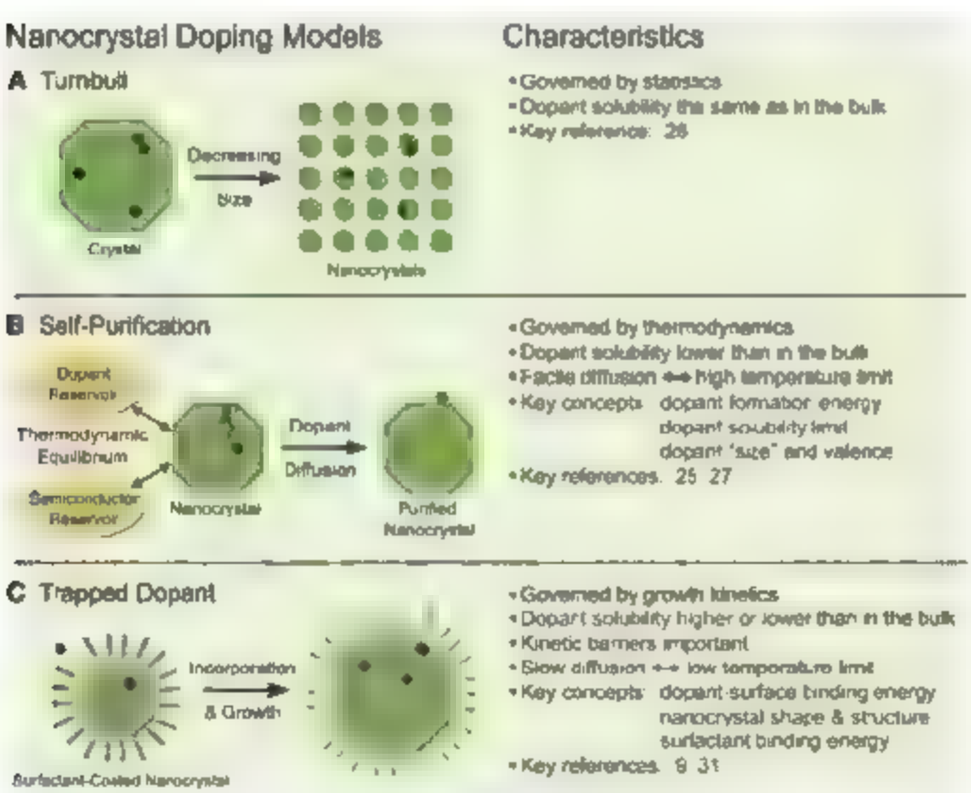


Fig. 2. Schematic and characteristics of three models used to explain doping in semiconductor nanocrystals. **(A)** Turnbull model, in which the number of impurities per nanocrystal decreases with crystallite size based solely on statistics (*26*). **(B)** Self-purification model, in which dopant solubility is lower in the nanocrystal than in the bulk semiconductor (*25*, *27*). **(C)** Trapped dopant model, in which kinetic factors govern the doping process (*9*, *31*).

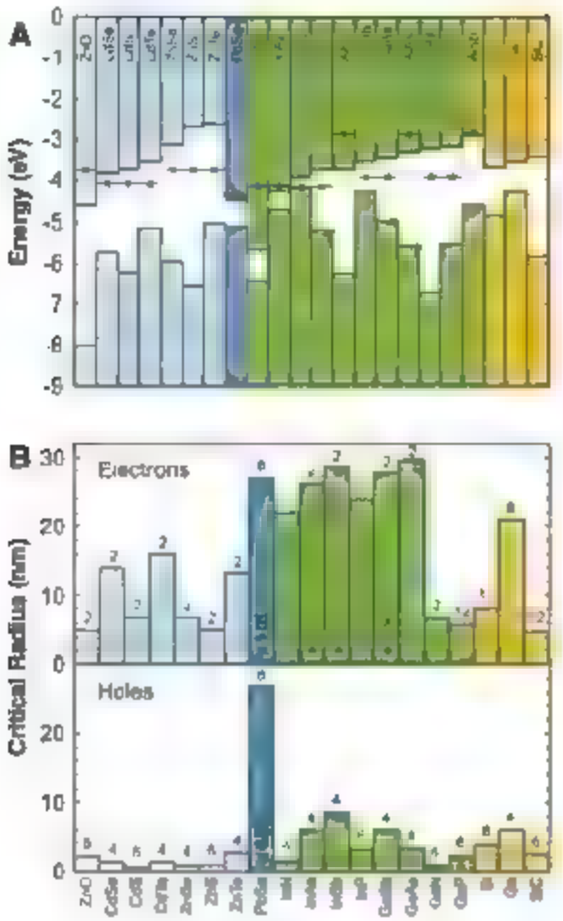


Fig. 3. (**A**) The energy of semiconductor electronic band edges relative to the vacuum level. The space between the bars represents the band gap between valence and conduction bands. Bulk values [courtesy of C. G. Van de Walle, University of California at Santa Barbara] are used (*38*), except for PbSe, which is estimated from nanocrystal results (*37*) and includes some confinement. Reduction potentials of most cations are marked with x's. From left to right, semiconductor groups are II-VI (pale blue), IV-VI (blue), III-V (green), and IV (gold). (**B**) Critical radii for autoionization of *n*- (top) and *p*-type (bottom) doped nanocrystals. When a nanocrystal has a radius smaller than this value, the impurity will automatically provide its carrier to the crystallite (*11*). A shallow hydrogen-like impurity level is assumed. Some radii have been multiplied by the scaling factors indicated. For each semiconductor, the degeneracies of the lowest-lying electron level (top) and hole level (bottom) are listed above each bar. Semiconductor groups are listed as in (**A**).

impurity solubility is much lower than in the bulk (*27*). This implies that nanocrystals are hard to dope for thermodynamic reasons and that impurities are expelled (Fig. 2B). However, a fundamental assumption in this view is that the nanocrystal is in thermodynamic equilibrium with its environment, including chemical equilibrium with reservoirs of impurity and nanocrystal constituents. The system then minimizes its Gibbs free energy. This principle underlies quantitative concepts such as the impurity formation energy and the solubility limit, as well as more heuristic ideas such as impurity size. For the assumption of thermodynamic equilibrium to be justified, however, atomic exchanges between the system and its reservoirs must be possible. This means that impurity atoms must be able to diffuse readily through the nanocrystal.

For growth above 1000°C in solid- or gas-phase syntheses, most dopants can diffuse rapidly. However, liquid-phase syntheses are performed typically below 350°C, where diffusion may or may not be facile. Impurities such as Ag and Au diffuse easily even at room temperature (*28*–*30*). Indeed, simply exposing CdSe crystallites to Ag leads to rapid, reversible transformations to Ag_2Se (*28*, *30*). In contrast, Mn diffusion in bulk II-VI crystals is negligible at liquid-phase growth temperatures (*31*) (although surface diffusion may still be facile).

If thermodynamic equilibrium is not established, then kinetic factors such as activation barriers will control doping (Fig. 2C). The impurity must adsorb on the surface of the nanocrystal and then be covered by additional material (*9*, *31*). For this to occur, the surface must be favorable for impurity binding. The nature of the surface depends on the crystal structure and nanocrystal shape, and hence these two factors affect doping. Moreover, in colloidal synthesis, surfactants must not bind to the surface or the impurity too strongly. Lastly, additional material must adhere so that, in the absence of diffusion, the impurity becomes trapped. These criteria have been used to explain Mn doping in II-VI nanocrystals (*31*).

Optical Properties

The fluorescence of the first Mn-doped nanocrystals was already more efficient than the best available undoped crystallites (*16*). The impurity localizes the excitation before nonradiative processes can occur. Researchers have since increased the fluorescence quantum yield of doped nanocrystals even further (to greater than 50%) by incorporating impurities at specific distances from the center of core-shell nanocrystals (*8*, *32*). Although the details are not yet understood, this result suggests that the competition between radiative and nonradiative pathways depends on the separation between the impurity, the surface, and other impurities (*8*, *20*, *32*). Placement of the dopant also influences the emission wavelength, presumably because of changes in its local environment. Figure 1C shows a series of ZnSe nanocrystals with Mn and Cu impurities. A range of colors is observed, in contrast to the ultraviolet emission from undoped ZnSe. These nanocrystals also exhibited higher resistance to photo-oxidation (*8*).

The influence of carriers on the optical properties was first studied through remote doping. Instead of using an impurity, researchers injected extra electrons into undoped nanocrystals with use of either a molecule attached to the nanocrystal surface (*33*) or an electrochemical cell (*34*). In these systems, the nanocrystal fluorescence was partially quenched. For example, Fig. 1D shows the green emission of undoped CdSe nanocrystals being controlled by an applied potential. This is consistent with the energy of any absorbed photons being transferred to the injected electron, which dissipates the energy as heat. In such experiments, new infrared transitions also appeared in the absorption spectra because of excitation of the added electron to higher-lying states. These optical results show that under certain circumstances the carrier can remain in the nanocrystal and avoid other fates, such as the electrochemical reactions discussed above.

Transport Properties

Remote doping experiments also revealed the influence of carriers on electrical properties. Although films of undoped CdSe nanocrystals are poor conductors, their conductivity increased by 12 orders of magnitude when the crystallites were brought into close contact and electrons were added (*35*). A similar increase was observed in PbSe nanocrystal films when hydrazine, an electron donor, replaced the surfactants initially bound to the PbSe surface (Fig. 1E) (*36*). Because transport in these films was also controllable by a gate voltage, impressive field-effect transistors could be demonstrated (Fig. 1F). In both CdSe and PbSe films, optical measurements revealed that the discrete electronic properties of the individual nanocrystals were still preserved.

Heterovalent impurity doping should lead to similar behavior in a broader range of nanocrystals. In colloidal nanocrystals, the most experimental progress has occurred in *n*-type impurity doping of ZnO. For example, extra electrons have been confirmed with Li impurities (*23*). This suggests that ZnO is somehow suited to retaining extra electrons. This conclusion is also consistent with remote doping experiments, where the lifetime of injected electrons in nanocrystals depended strongly on the host semiconductor, decreasing from ZnO to CdSe to ZnSe. In addition, the lifetime decreased with nanocrystal size.

Guyot-Sionnest *et al.* have explained this behavior in terms of the energy of the extra electron (*33*, *34*, *37*). Figure 3A shows the position of the valence and conduction bands for several semiconductors relative to the vacuum level (*38*). In the bulk, an extra electron would sit at the bottom of the conduction band. This energy increases from ZnO to CdSe to ZnSe. In nanocrystals, it is increased further by confinement, which increases with decreasing size. Thus, in moving from ZnO to CdSe to ZnSe or to small nanocrystals, the extra electron becomes less

stable relative to competing electrochemical reactions (*39*). For example, the energy of an electron in ZnSe can exceed the reduction potential of the cation, indicated in Fig. 3A for comparison. If this reaction occurs ($Zn^{2+} + 2e^- \rightarrow Zn$), the electron is consumed, and the crystallite can lose material through corrosion. Similarly, an extra hole in a bulk crystal will sit at the top of the valence band. In a nanocrystal, the hole will be pushed lower in Fig. 3A by confinement. The lower the hole, the more likely electrochemical reactions become.

This description qualitatively explains observed stability trends. However, it neglects details such as the nanocrystal surroundings and kinetic barriers to reactions. For example, the reduction potentials in Fig. 3 assume aqueous conditions, which are atypical for colloidal nanocrystals. The reduction potentials in organic solvents shift higher, making reactions less favorable and increasing carrier stability. Figure 3A thus represents a lower bound for stability. Nevertheless, it can guide the selection of semiconductor hosts for obtaining stable carriers.

Talapin and Murray identified two other factors to explain why remotely doped films of PbSe nanocrystals were more conductive than those of CdSe (*36*). First, a carrier in a PbSe nanocrystal is more strongly confined because of the larger dielectric constant. Thus, in PbSe the carrier leaks more into the crystallite surroundings, which increases conductivity. For impurity doping, this can be quantified by a critical radius for auto-ionization (Fig. 3B). A larger critical radius is desirable. Second, the lowest-lying conduction state in PbSe nanocrystals is highly degenerate, indicating that many channels exist for charge transport. To increase conductivity, a semiconductor with high degeneracy in the electron or hole states (Fig. 3B) should be chosen.

In the remote doping experiments, some of the injected electrons were consumed by unidentified traps before the nanocrystal states were occupied, even when the carriers should be stable (Fig. 3A). Although the origin and number of these traps is debated, for impurity doping it would be detrimental if they were to outnumber the carriers provided by the impurities.

Unconventional Approaches

A completely different approach has very recently been introduced to obtain nanocrystals with carriers (*40*). Conductivity was observed in a thin film of PbTe and Ag_2Te nanocrystals that self-organized into a periodic crystal of nanocrystals (Fig. 1G), or superlattice (*3*). Although the mechanism that leads to conductivity is not understood, this result suggests that Ag_2Te nanocrystals provide carriers to the film. Future efforts may explore the use of these nanocrystal "impurities" in solids of nanocrystals to create a broad range of materials to explore.

Outlook

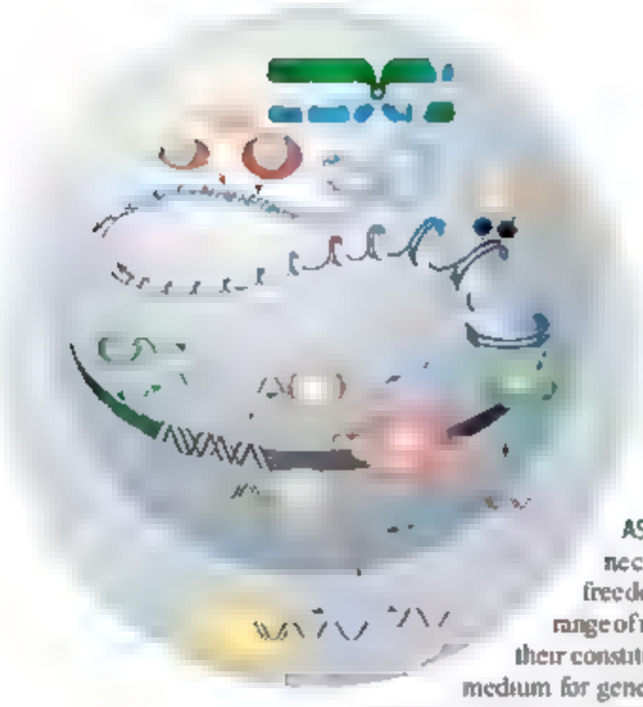
Doping enhances the properties of nanocrystals by providing another means to control their remarkable electronic, optical, transport, and magnetic properties. In this sense, the development of doped nanocrystalline materials is mirroring the evolution of bulk semiconductors half a century ago. Future progress in doped nanocrystals will require improving synthetic control over dopant incorporation, optimizing their concentrations, and investigating the phenomena that emerge. Pursuing these goals and applications will be an exciting and challenging task.

References and Notes

1. A. P. Alivisatos, *Science* **271**, 933 (1996).
2. Y. Yin, A. P. Alivisatos, *Nature* **437**, 664 (2005).
3. C. B. Murray, C. R. Kagan, M. G. Bawendi, *Annu. Rev. Mater. Sci.* **30**, 545 (2000).
4. V. I. Klimov *et al.*, *Nature* **447**, 441 (2007).
5. X. Michalet *et al.*, *Science* **307**, 538 (2005).
6. L. Gur, N. A. Fromer, M. L. Geier, A. P. Alivisatos, *Science* **310**, 462 (2005).
7. C. Wang, B. L. Wehrenberg, C. Y. Woo, P. Guyot-Sionnest, *J. Phys. Chem. B* **108**, 9027 (2004).
8. N. Pradhan, D. Goorskey, J. Thessing, X. G. Peng, *J. Am. Chem. Soc.* **127**, 17586 (2005).
9. J. D. Bryan, D. R. Gamelin, *Prog. Inorg. Chem.* **54**, 47 (2005).
10. M. L. Efros, E. L. Rashba, M. Rosen, *Phys. Rev. Lett.* **87**, [illegible] (2001).
11. A. I. Ekimov, I. A. Kudryavtsev, M. G. Ivanov, Al. L. Efros, *J. Lumin.* **46**, 83 (1990).
12. A. Henglein, *Ber. Bunsenges. Phys. Chem.* **86**, 301 (1982).
13. L. E. Brus, *J. Chem. Phys.* **79**, 5566 (1983).
14. M. T. Swihart, *Curr. Opin. Colloid Interface Sci.* **8**, 127 (2003).
15. T. A. Kennedy, E. R. Glaser, P. B. Klein, R. N. Bhargava, *Phys. Rev. B* **52**, R14356 (1995).
16. R. N. Bhargava, D. Gallagher, X. Hong, A. Nurmikko, *Phys. Rev. Lett.* **72**, 416 (1994).
17. D. M. Hoffman *et al.*, *Solid State Commun.* **114**, 547 (2000).
18. D. J. Norris, N. Yao, F. T. Charnock, T. A. Kennedy, *Nano Lett.* **1**, 3 (2001).
19. J. F. Suyver, S. F. Wuister, J. J. Kelly, A. Meijerink, *Phys. Chem. Chem. Phys.* **2**, 5445 (2000).
20. P. V. Radovanovic, D. R. Gamelin, *J. Am. Chem. Soc.* **123**, 12207 (2001).
21. K. M. Hanif, R. W. Meulenberg, G. F. Strouse, *J. Am. Chem. Soc.* **124**, 11495 (2002).
22. C. A. Stowell, R. J. Wiacek, A. E. Saunders, B. A. Korgel, *Nano Lett.* **3**, 1441 (2003).
23. S. B. Orlinskii *et al.*, *Phys. Rev. Lett.* **92**, 047603 (2004).
24. M. Fujii, Y. Yamaguchi, Y. Takase, K. Ninomiya, S. Hayashi, *Appl. Phys. Lett.* **87**, 211919 (2005).
25. F. V. Mikulec *et al.*, *J. Am. Chem. Soc.* **122**, 2532 (2000).
26. D. Turnbull, *J. Appl. Phys.* **21**, 1022 (1950).
27. G. M. Dalpian, J. R. Chelikowsky, *Phys. Rev. Lett.* **96**, 226802 (2006).
28. D. H. Son, S. M. Hughes, Y. Yin, A. P. Alivisatos, *Science* **306**, 1009 (2004).
29. T. Mokari, A. Aharoni, I. Popov, U. Banin, *Angew. Chem. Int. Ed.* **45**, 8001 (2006).
30. R. D. Robinson *et al.*, *Science* **317**, 355 (2007).
31. S. C. Erwin *et al.*, *Nature* **436**, 91 (2005).
32. Y. Yang, O. Chen, A. Angerhofer, Y. C. Cao, *J. Am. Chem. Soc.* **128**, 12428 (2006).
33. M. Shim, P. Guyot-Sionnest, *Nature* **407**, 981 (2000).
34. C. Wang, M. Shim, P. Guyot-Sionnest, *Science* **291**, 2390 (2001).
35. D. Yu, C. J. Wang, P. Guyot-Sionnest, *Science* **300**, 1277 (2003).
36. D. V. Talapin, C. B. Murray, *Science* **310**, 86 (2005).
37. B. L. Wehrenberg, P. Guyot-Sionnest, *J. Am. Chem. Soc.* **125**, 7806 (2003).
38. C. G. Van de Walle, J. Neugebauer, *Nature* **423**, 626 (2003).
39. R. J. D. Miller, G. L. McLendon, A. J. Nozik, W. Schmickler, F. Willig, *Surface Electron Transfer Processes* (VCH, New York, 1995).
40. J. J. Urban, D. V. Talapin, E. V. Shevchenko, C. R. Kagan, C. B. Murray, *Nat. Mater.* **6**, 115 (2007).

41. We thank P. Alivisatos, M. Bawendi, M. Brandt, P. Guyot-Sionnest, A. Meijerink, M. Stuttmann, and D. Vanmaekelbergh for stimulating discussions and T. Kennedy, A. Wills, and L. Zu for help with the figures. This work was supported in part by the Office of Naval Research and the NSF Nanoscale Interdisciplinary Research Team program (CBET-0506672 and CBET-0506748). D.J.N. benefited from a fellowship from the Alexander von Humboldt Foundation.

10.1126/science.1143802



INTRODUCTION

Freedom of Expression

AS IN CIVIL SOCIETY, WHERE THERE MUST necessarily be checks and balances on freedom of expression, cells have evolved a range of mechanisms to regulate the expression of their constituent genes. By far the best-understood medium for gene regulation is the protein transcription factor. The broad set of rules by which these regulators operate is outlined by Hobert (p. 1785). However, new and unexpected gene regulatory systems have been discovered in the past decade, perhaps the most important of which involve microRNAs (miRNAs). Hobert compares the action of these small noncoding RNAs, found in many eukaryotes, with their proteinaceous counterparts, showing that miRNAs share many similar activities but also display unique traits in their compartmentalization, rapid reversibility, and evolvability. Makeyev and Maniatis (p. 1789) provide examples of the profound systemwide influence that miRNAs can have on gene expression programs. miRNAs are also being linked to a growing list of common ailments, including cancer, heart disease, diabetes, and viral illnesses such as hepatitis. In a related News story (p. 1782), Jennifer Couzin explores how miRNAs are attracting the interest of biomedical researchers and biotechnology companies eager for new ways to diagnose and treat diseases.

Another recently discovered RNA-based regulatory system is the riboswitch, found in plant, fungal, and prokaryotic RNAs. Although they possess a deceptively simple bipartite structure, Breaker (p. 1795) describes how their chemistry, conformation, and kinetics have facilitated the evolution of sophisticated gene-control systems. Indeed, the overwhelming regulatory potential of RNA is graphically described by Amaral *et al.* (p. 1787), who list the many and varied instances in which RNA has been implicated in regulatory events.

This is not to suggest that research on transcription factors is moribund—far from it, as revealed by Core and Lis (p. 1791), for example, who discuss the revival of earlier work revealing a critical regulatory step, the pausing of the RNA polymerase II molecule, during the early phase of transcription elongation. The often highly dispersed nature of transcription factor binding sites in many eukaryotic genes provided the first clues that the spatial organization of the genome can be critical for gene regulation; for example, allowing combinatorial interactions between genes and regulatory elements, as described by Dekker (p. 1793). Understanding the origins of these regulatory systems requires that we examine how they have evolved, prompting Tuch *et al.* (p. 1797) to note that orthologous regulatory circuits with similar transcriptional outputs can nonetheless undergo massive rewiring in even closely related species.

Several gene regulatory systems are also highlighted in our online sister journal *Science* Signaling (www.sciencemag.org/cgi/content/full/sigtrans/1/19/eg1). how oncogenic *Ras* causes the epigenetic silencing of *Fas* and other tumor-suppressor genes, how intrachromosomal looping positions enhancers close to the promoter of the tumor necrosis factor–α gene to stimulate its expression in activated T cells, and how the abundance of the transcriptional coactivator steroid receptor coactivator–3 controls estrogen-dependent gene transcription.

–GUY RIDDIHOUGH, BEVERLY A. PURNELL, JOHN TRAVIS

Gene Regulation

CONTENTS

See also related Editorial page 1733, online material page 1727 or at www.sciencemag.org/cgi/content/full/319/5870/1781

Science

CREDIT: P. P. AMARAL ET AL.

NEWS

MicroRNAs Make Big Impression In Disease After Disease

Hunting for new ways to diagnose and treat common diseases, biologists and companies are racing to decipher the promise of these RNAs

When it comes to RNA molecules, the deeper biologists dig the more they seem to uncover. The 2006 Nobel Prize honored the discovery of RNA interference, in which scientists use short strands of the chemical to silence specific genes, and RNAi has helped shed light on the fact that cells naturally use RNA molecules just 20 to 22 nucleotides long, dubbed microRNAs, to regulate gene expression. Lately, microRNAs are garnering attention on the biomedical front, startling researchers with ever-expanding roles in disease. A flood of studies show that microRNAs may offer a window into the development of various ailments, including cancer, diabetes, and heart failure, and provide a chance to strike disease targets that until now were unreachable.

"People are essentially stunned that this whole level of regulation existed, and we just didn't know about it until a few years back," says developmental biologist Frank Slack of Yale University, who has branched out from studying microRNAs in cancer to linking them to Alzheimer's disease and life span. When poorly regulated, the molecules appear to drive cancer and a host of other diseases, and this could make them useful in diagnosing disease very early. MicroRNAs made by viruses, meanwhile, may help pathogens gain a foothold in their host, which could suggest new targets for antiviral drugs.

Biologists who have made some of the early microRNA discoveries are eager to push toward new treatments for patients, hoping to also cash in, many are teaming up with biotechnology companies or establishing their own. Companies that formed several years ago to capitalize on different types of RNA molecules, notably ones used for RNAi or others known as antisense, are now expanding into the microRNA arena.

Bringing MicroRNA Discoveries to the Clinic

Company	Location	Disease Focus	Founded
Asuragen	Austin, Texas	Cancer diagnostics	2006
Crogen Pharmaceuticals	Philadelphia, Pennsylvania	Cancer	2004
Miragen Therapeutics	Boulder, Colorado	Cardiovascular and muscle diseases	2007
Regulus Therapeutics	Carlsbad, California	Viral diseases, cancer	2007
Rosetta Genomics	Rehovot, Israel, and Jersey City, New Jersey	Cancer	2000

New view of the body. From the heart to the blood to the pancreas and beyond, scientists are finding tantalizing hints that microRNAs can help keep us healthy or make us sick. Biotech firms are springing up to convert these discoveries into new products that can diagnose, treat, or predict the course of disease.

Two microRNA-based therapeutic strategies are being considered: delivering mimics of the molecules that could promote health or blunting the impact of ones that contribute to disease. As with any novel therapy, microRNAs come with their own set of challenges that must be overcome before testing begins in people. Delivering the molecules to the right cells is still a technological hurdle, as is identifying the genes microRNAs influence to ensure that modifying their expression won't have untoward effects. Many individual microRNAs home in on dozens or even hundreds of genes.

Adding a reason for caution, biologists conducting animal tests have been taken aback by the dramatic effects of slightly dialing up or down the dose of a single microRNA. Many believe microRNAs will lead to therapeutics but warn that it will take time. "The microRNA network is really subtle," says geneticist John Rossi of City of Hope in Duarte, California, who is considering how microRNAs might modulate HIV infections as well as behavioral syndromes such as schizophrenia. He predicts that many correlations between microRNAs and disease "are going to be wrong in the end. But you have to start somewhere."

Help for faltering hearts

Out of the 25,000 or so human genes, scientists have identified about 500 that yield microRNAs, and they continue to surprise. Consider an experiment that caused Deepak Srivastava, who directs the Gladstone Institute of Cardiovascular Diseases at the University of California, San Francisco, to do a double take. Srivastava studies cardiac development and gene pathways that, when disrupted, can lead to congenital or adult heart disease. When he learned of microRNAs, which were first discovered in 1993 in worms, he began hunting for ones expressed at high or low levels in the heart, first in mice and in fruit flies and later in people. After identifying about 10 such microRNAs, he pursued them one by one, testing whether changing expression levels led to changes in heart function or development. Last April, he and his colleagues described in *Cell* one of the first examples of mice engineered to lack a specific microRNA, called miR-1-2.

Although miR-1-2 is highly expressed in heart muscle, Srivastava's team wasn't expecting their mice to look much different from normal ones. That's because miR-1-2 has a twin, an identical DNA sequence on another chromosome that the researchers didn't delete. By knocking out one copy, they dialed down the dose of this microRNA by 50%. The effect was dramatic: Half the animals died of holes in the heart, and others were found to have fatal disruptions in cardiac rhythm (*Science*, 27 April 2007, p. 530).

"At the time, it was a really surprising result," says Srivastava. Today, he believes, it "makes a little more sense." That's because scientists are now learning that microRNAs are "the regulators of the master regulators," as Srivastava puts it—potentially the controllers of entire pathways of genes. The cardiac defects seen in these mice resemble those that are

among "the most common heart defects in humans," says Srivastava, although no one has proven yet that the human defects are caused by a microRNA deficiency, too. Srivastava, who's considering helping establish a company, is now studying how microRNAs govern the development of the heart's four chambers, each of which displays a unique gene-expression pattern and has a distinct function.

Although Srivastava's work hints that problems with microRNAs could explain some heart defects seen at birth, these bits of nucleic acids may also affect the adult heart. At the University of Texas Southwestern Medical Center in Dallas, molecular biologist Eric Olson and his postdoctoral fellow Eva van Rooij are using mice to connect microRNAs to common heart diseases such as cardiac hypertrophy, in which the heart's walls thicken and the organ struggles to keep pumping. Eventually, the condition leads to heart failure; there are few effective treatments. The pair has identified a mouse microRNA, miR-208, whose DNA sequence is hidden within a gene that encodes the muscle protein myosin, which helps the heart contract and function under stress. More recently, they've found that many myosin genes contain microRNAs that regulate one another to keep heart muscle healthy—an interconnection among microRNAs that no other group has described, says van Rooij.

By manipulating microRNAs such as miR-208, "we can really start attacking disease in a whole different way," predicts William Marshall, a biologist and chemist in Boulder, Colorado. Marshall, who has worked at various RNA-focused biotechnology companies, met Olson through a mutual friend. They recruited others to launch the company Miragen Therapeutics last August. Marshall is the president and chief executive officer. Van Rooij will be the director of research beginning in January 2009.

The cardiology field has an advantage over others: Its doctors have experience supplying drugs straight into the target organ, for example, by injection into the coronary arteries. It should be doable, some predict, to shoot extra microRNAs, or microRNA suppressors, directly into the heart. "I think we might see the first trials [of a microRNA-based therapy] in the cardiology field," says Markus Stoffel, a molecular biologist at the Swiss Federal Institute of Technology in Zürich.

Target practice

Researchers such as Stoffel are hoping that microRNAs won't share the woes of a related technology that disappointed back in the 1990s.

CREDIT: THE ROCKEFELLER UNIVERSITY

MicroRNA hunter. Molecular biologist Markus Stoffel is chasing early signs of a connection between the small RNAs and diabetes.

called antisense therapy. Like potential microRNA therapies, antisense sought to modulate gene expression but did so by targeting the messenger RNAs that translate genes into proteins, as a way of blocking protein synthesis. Researchers found that antisense "just doesn't work very well for inhibiting messenger RNAs," says geneticist and molecular biologist Joshua Mendell of Johns Hopkins University in Baltimore, Maryland, who focuses on microRNAs and cancer. Only one antisense drug has received approval, in 1998.

In some diseases, microRNAs are overabundant, and the goal will be to dial down their expression by injecting a complementary RNA sequence that binds to and disables the target microRNA. But in other ailments, such as certain cancers, microRNAs appear in lower concentrations than in normal tissue, suggesting that treatments will need to add microRNA "mimics."

> *"People are essentially stunned that this whole level of regulation existed, and we just didn't know about it until a few years back."*
>
> —FRANK SLACK, YALE UNIVERSITY

For the overabundance problem—which requires blunting microRNA expression—many scientists are now using a strategy designed by Stoffel in 2005. Stoffel studies microRNAs in diabetes and metabolism and realized that he needed a delivery system to get RNA sequences into cells where they could silence microRNAs. To accomplish this, Stoffel collaborated with Muthiah Manoharan of Alnylam Pharmaceuticals, an RNAi company in Cambridge, Massachusetts, to create "antagomirs," so called because they antagonize the miR, or microRNA. Stoffel's antagomirs are RNA snippets linked to cholesterol molecules, which help slip the silencers into cells. After being injected into the tail veins of mice, antagomirs travel through the body, they have successfully modified microRNA expression in many organs. Antagomirs can't cross the blood-brain barrier, but scientists have injected them directly into the brain, where they penetrated brain cells. By the third day after an antagomir injection, Stoffel says, the microRNAs targeted disappear and stay silent for weeks. He believes that's because the antagomirs remain in the cells for some time, blunting any new microRNAs a cell produces.

Alnylam, of whose scientific advisory board Stoffel is a member, has acquired rights to the antagomirs. And in September, Alnylam and another company, Isis Pharmaceuticals, pooled their intellectual property to create a new microRNA-focused company called Regulus Therapeutics.

It's not clear yet whether antagomirs will remain the delivery vehicle of choice for microRNA silencers because the cholesterol they contain might harm the liver, says Srivastava. Although he uses them in animal experiments without apparent side effects, Srivastava suspects that they may not be acceptable for treating people.

The mirror image of these efforts—the attempt to boost rather than silence microRNAs for therapeutic purposes—may be just as big a challenge. When it comes to overexpressing a microRNA, "it's not as straightforward as I thought it would be," says Stoffel. The favored approach involves supplementing a weak supply of a microRNA by administering synthetic precursors, which cells would take up and process into mature microRNAs. But how to get the precursors into the body's cells isn't clear. Developing traditional small-molecule drugs that boost the activity of genes that encode microRNAs is widely considered a long shot. "I don't hold out a lot of hope for it," says Marshall.

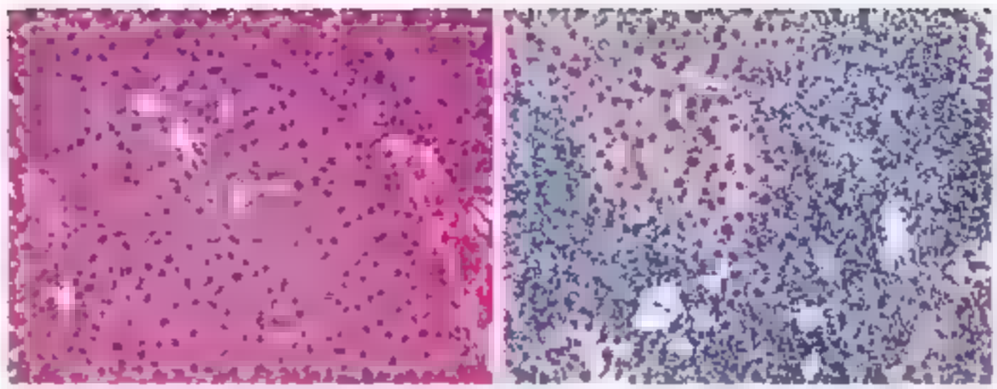
From the pancreas to the liver

Stoffel has juggled his work on antagomirs with another pet project, identifying microRNAs that may play a role in diabetes.

Stoffel screened insulin-producing beta cells derived from a mouse pancreas for highly expressed microRNAs. Ten showed up that had not been identified in other organs, and Stoffel's group has focused on miR-375, the one most strongly expressed by the beta cells. Unpublished work Stoffel has done in mice suggests that miR-375 helps islets in the pancreas adapt to certain stresses, such as pregnancy or obesity, which call on the body to produce extra insulin.

MicroRNA researchers are also considering how to beat back viral infections. Molecular virologist Bryan Cullen of Duke University in Durham, North Carolina, has puzzled over how viruses invading the human body interact with natural RNAi machinery. But recently, he began focusing on the microRNAs made by some viruses and whether they might give the invading pathogens an advantage against their host. Many viruses in the herpesvirus family, for example, "make a ton of microRNAs" themselves. "The champion at this point is Epstein-Barr virus, with 23," he says.

Preliminary research by labs in France and Germany suggests that disabling some of these viral microRNA genes makes the viruses less harmful to mice. But at this point, "it's hard for us to tell what's going on," says Cullen. He is collaborating with the company Regulus to determine just how important microRNAs are for the viruses that carry them. And viruses also appear to exploit certain human microRNAs: Regulus's first project will be blunting a microRNA in the human liver that seems to help the hepatitis C virus replicate, says John Maraganore, Alnylam's CEO.



Fast track to cancer. Overexpressing a particular microRNA in some of a mouse's immune cells leads to leukemia; in a mouse without extra microRNA (*left*), liver tissue (pink) is normal, but in one with boosted microRNA (*right*), leukemia cells (blue) infiltrate the liver.

CREDIT: CARLO CROCE AND STEFAN COSTINEAN

Detecting cancer's first steps

The starting point for microRNAs' role in disease was cancer, and that line of inquiry remains arguably the most active. Broad screens of human tumor tissue have shown that microRNAs tend to be expressed differently in cancer cells compared with normal tissue of the same organ. Moreover, patterns of microRNA expression—certain ones overexpressed, others underexpressed—correlate with disease prognosis, according to retrospective studies of cancer patients.

In some cases in which a microRNA is underexpressed in a cancer, replenishing its supply in cancerous cells stops the disease from proliferating in animal models. But at least in petri dishes, extra amounts of certain microRNAs don't appear to affect normal cells, which already boast abundant supplies.

In mice, certain microRNAs can be manipulated to drive cancer. Carlo Croce of Ohio State University, Columbus, one of the first in the cancer-microRNA arena, and others have found that artificially upregulating or downregulating particular microRNAs can initiate or spur on the disease. Croce has founded a company, Crogen Pharmaceuticals, to tackle microRNA-based diagnostics, prognostics, and therapeutics in cancer. One of the first microRNA companies, Rosetta Genomics, was launched in Israel in 2000 and is also pouring resources into cancer.

> *"We can really start attacking disease in a whole different way."*
>
> —WILLIAM MARSHALL, MIRAGEN THERAPEUTICS

To nail down the role of microRNAs in tumors, it's crucial to develop genetically engineered animals born without specific microRNAs, says Mendell of Johns Hopkins. "We need to do a better job of documenting their roles in cancer and documenting the mechanisms" by which they act.

That's tougher than it sounds. For example, Mendell has found that the protein produced by the oncogene *myc*, which is frequently active in cancer cells, downregulates dozens of microRNAs. But that's far from the whole story, for the microRNA-gene network is unimaginably complex. Although some proteins made by oncogenes home in on microRNAs, as Mendell describes, the reverse is also true, with other microRNAs controlling the activity of oncogenes. "The result is a series of interactions that can have a very potent effect" on cancer, says Mendell.

Beyond exploiting microRNA biology to treat cancer, many are eyeing potential diagnostics that would detect cancer at an early stage or diagnose it when standard approaches fail. One strategy is to examine microRNA expression patterns in metastatic tumors of unknown origin, a problem that occurs in 2% to 4% of cancers and presents treatment challenges. In 2005, a team led by Todd Golub at the Dana-Farber Cancer Institute in Boston described in *Nature* their effort to classify 17 tumors that couldn't be cataloged based on their appearance. They correctly identified the origins of 12 of the 17 based on microRNA patterns in tumors that vary slightly depending on where in the body the tumor originated. Traditional gene-expression signatures using messenger RNAs correctly identified just one.

Rosetta Genomics officials say they hope to begin selling three microRNA-based cancer diagnostic tests later this year; one will pin down tumors of unknown origin, and the others will help doctors distinguish between different lung-related cancers. In 2006, the company Asuragen in Austin, Texas, was launched and now focuses on microRNAs in body fluids as a cancer-diagnosis tool. Slack of Yale is one of the academics it's collaborating with. Company officials won't say much about their plans, beyond saying that Asuragen will concentrate on many major cancers, including lung, prostate, colon, breast, and stomach, and expects the development of microRNA-based diagnostics to take several years. Many microRNAs overlap among a number of cancers, which makes commercializing them less daunting. Recently, Asuragen closed on a second round of funding, garnering $18.5 million.

MicroRNAs in medicine "is a hot field," agrees Bruce Booth, a venture capitalist at Atlas Venture in Waltham, Massachusetts, who helped found and fund Miragen. Still, no one knows whether targeting microRNAs will wind up helping people—but if they do, they may play a huge role in patient care. "We're willing to take on higher risk with such significant upsides," says Booth. Like many others, he's betting big that his investment will pay off.

—JENNIFER COUZIN

PERSPECTIVE

Gene Regulation by Transcription Factors and MicroRNAs

Oliver Hobert

The properties of a cell are determined by the genetic information encoded in its genome. Understanding how such information is differentially and dynamically retrieved to define distinct cell types and cellular states is a major challenge facing molecular biology. Gene regulatory factors that control the expression of genomic information come in a variety of flavors, with transcription factors and microRNAs representing the most numerous gene regulatory factors in multicellular genomes. Here, I review common principles of transcription factor– and microRNA-mediated gene regulatory events and discuss conceptual differences in how these factors control gene expression.

Transcription factors (TFs) and microRNAs (miRNAs), the largest families of trans-acting, gene regulatory molecules in multicellular organisms, share a common regulatory logic (*1*) (Fig. 1 and figs. S1 and S2). Sets of combinationally expressed TFs ("TF codes") and miRNAs ("miRNA codes") precisely delineate individual cell types. By binding to discrete cis-regulatory elements, individual TFs and miRNAs can control dozens, if not hundreds, of target genes. Moreover, most, if not all, genes in the genome are controlled not by a single, but by a combination of trans-acting factors (Fig. 1). Many TFs bind cooperatively to their cognate DNA sequences and/or cooperatively recruit transcriptional cofactors (*2*, *3*). Similarly, cooperative action of miRNAs has also been defined through reporter gene assays (*4*). Cooperativity therefore provides the mechanistic basis for reading out combinatorial expression patterns of both TFs and miRNAs (Fig. 1).

Binding site accessibility provides an additional layer of gene regulatory control (Fig. 1). In vivo occupancy of TF binding sites depends on nucleosome coverage of the site, with nucleosome positioning and remodeling being regulated processes (*5*). Similarly, the accessibility of a miRNA recognition site is controlled by a member of a large (>100) family of RRM domain containing RNA binding proteins (*6*). Other family members may control other mRNA/target interactions. Apart from protein-regulated site accessibility, miRNA binding site accessibility is also controlled by folding of the mRNA target sequences into secondary structures (*7*). Mere coexpression of a miRNA and its mRNA target, therefore, does not always lead to a functional interaction—a notion to be kept in mind when considering computer-predicted miRNA/target interactions.

The Importance of Repression

Whereas TFs are known to positively or negatively regulate transcription, miRNAs appear to regulate gene expression mostly, but not always, through repression (*8*, *9*). Gene repression is an important mechanism to shape cell-specific gene regulatory programs. Broad and non-cell type specific transcriptional activation events, evoked by broadly expressed TFs, gain specificity through the action of cell type specific transcriptional repressors, which restrict gene expression to a smaller subset of cells (*1*). Moreover, in many developmental contexts, activating effects of TFs often turn out to be double-negative, derepression effects in which a TF "activates" transcription by repressing expression of a transcriptional repressor (*1*, *10*). The repressive mode of miRNA action therefore fits neatly into the overall importance of gene repression in defining cell-specific gene expression programs.

Regulating Regulators—Networks and Modifications

The ability of TFs to define, either alone or in combination, cell type–specific gene expression programs rests on cell type–specific expression profiles of the TFs themselves. Such profiles are controlled by more upstream layers of gene regulatory programs. Developmental processes can therefore be considered as a succession of hierarchically acting regulatory states (*11*). Consequently, transcriptional regulatory programs have been placed into well-defined regulatory networks that are characterized by specific, small network motifs such as positive and negative feedback and feedforward motifs that endow the system with specific properties such as signal amplification, dampening, persistence detection, and oscillation (*12*). Because the expression of many TFs themselves is subject to miRNA regulation and the cell type–specific expression profiles of miRNAs are brought about largely by conventional TF-dependent transcriptional control mechanisms, it does not come as a surprise that miRNAs and TFs are linked to one another in gene regulatory networks (*13*, *14*). As with TF-only networks, prevalent motifs within these networks include feedforward and feedback loops (Fig. 1 and fig. S2).

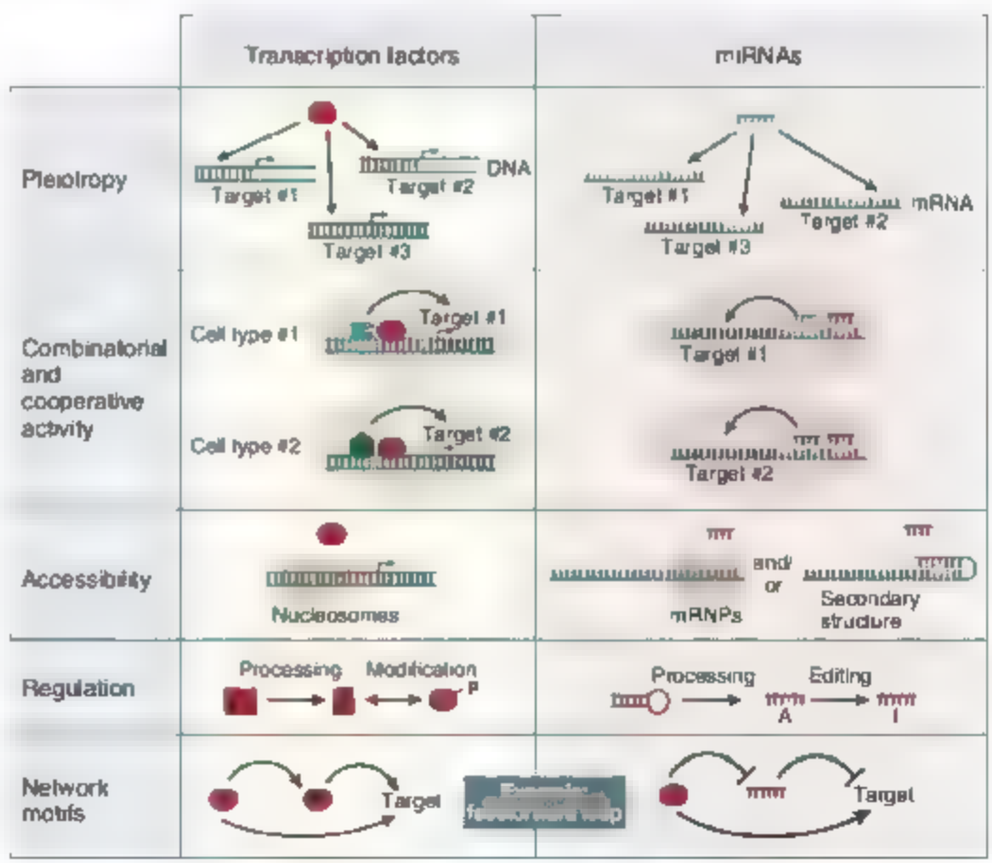


Fig. 1. A schematic visualization of some shared principles of TF and miRNA action. See also fig. S2.

Howard Hughes Medical Institute, Department of Biochemistry and Molecular Biophysics, Columbia University Medical Center, New York, NY 10032, USA. E-mail: or38@columbia.edu

Besides being regulated at the level of gene expression, TF activity is prominently regulated via posttranslational events such as protein phosphorylation, processing, or localization. miRNA-mediated control of gene expression may also be subject to posttranscriptional regulation (Fig. 1). For example, the processing of miRNAs from their primary transcript to their mature form is subject to regulation in various cell types (*15*, *16*). Modifications of miRNA species by RNA editing also control miRNA activity in a cell type–specific manner (*17*). Posttranslational modifications of additional protein factors involved in miRNA function may introduce an additional layer of regulatory complexity in miRNA function.

Phenotypic Spectrum

The points raised above provide a somewhat biased view in which many features of miRNA- and TF-mediated regulatory events appear quite similar (Fig. 1). But are miRNA and TFs really on par in terms of their importance for gene regulatory events? Genetic analysis in unicellular and multicellular organisms has amply demonstrated the importance of TFs in controlling development and homeostasis (*11*). Elimination of specific animal miRNAs can also produce striking phenotypes in multiple patterning events (*18*). However, the deletion of >80% of individual miRNA loci in *Caenorhabditis elegans* revealed that less than 10% of miRNA knockouts result in clear developmental or morphological defects (*19*). In contrast, RNA interference–mediated loss-of-function analysis shows that about 30% of all *C. elegans* TF losses cause easily observable phenotypes (*20*). Redundancy between closely related miRNA family members may mask essential functions (*19*), but as exemplified by the miRNA *lsy-6*, which diversifies the functional capacities of two very similar chemosensory neurons (*21*), miRNA use may be more biased toward controlling specific aspects of terminal differentiation programs of individual cell types. In support of this notion, miRNAs tend to have highly cell type specific expression profiles, both during normal development and in specific disease states, such as cancer (*22*, *23*). Are there intrinsic features of miRNAs that would explain why they may cover a more restricted regulatory niche as compared to TFs?

Unique Features of miRNA-Mediated Gene Regulation

The gene regulatory region of TF targets is often complex and can span dozens of kilobases, whereas miRNA-controlled 3′-untranslated regions are, on average, <1 kb in size (*11*). Not only does this limit the amount of regulatory inputs that a gene can sample from miRNAs versus TFs, but it also provides less substrate for evolution to "play on," i.e., to evolve new regulatory inputs, a major driving force in evolution (*24*). This inherent size restriction may disfavor the evolution of miRNA-dependent gene expression programs. On the other hand, inherent features of miRNA-mediated gene regulation may also provide unique evolutionary opportunities. Whereas, in principle, expression of a gene can become more restricted on an evolutionary time scale by acquiring a novel repressive TF binding site in its promoter, such an acquisition may be constrained by features of the transcriptional activation program. For example, evolving a negative regulatory input into a promoter that is combinatorially activated by several distinct transcriptional activators may be mechanistically more difficult—and evolutionarily more improbable than evolving a miRNA input into the mRNA product of the transcriptional event. By acquiring a miRNA binding site or evolving a novel miRNA gene de novo that matches a given mRNA, evolution is essentially provided with an independent opportunity to diversify a gene expression program. In sum, the fact that miRNAs can only serve to further modify transcriptional programs because they require an mRNA substrate to work on—and the more restricted size of their regulatory control regions may explain the more restricted phenotypic spectrum of miRNAs compared to TFs.

Speed and reversibility are other distinguishing features of miRNA-mediated gene regulation that may result in a more specialized regulatory niche of miRNAs. For transcription to be repressed, a sophisticated machinery needs to be set in place in a subcellular compartment—the nucleus—that is distinct from the production site of the protein product, the cytoplasm. The stability of already transcribed mRNA species sets another limit to the speed with which transcriptional repression can wipe out the expression of a target gene. In contrast, miRNAs can rapidly turn off protein production right at the site of protein production, the ribosome. Another factor that speeds up miRNA-mediated control of gene expression is that, owing to their small size and noncoding nature, miRNAs may be produced more rapidly than TFs, thereby decreasing response times to stimuli that induce gene repression. Lastly, a miRNA-repressed target can also be reactivated more rapidly than a transcriptionally repressed target, because such reactivation may merely involve the translocation of an already present mRNA to an active ribosome (*25*–*27*).

Another important conceptual difference between miRNA- and TF-mediated gene regulation is that miRNA action can be compartmentalized within a cell to rapidly alter gene expression locally. For example, highly compartmentalized neurons face the challenge of translationally regulating gene expression on a synapse-specific, rather than cell-wide, level (*28*). TFs cannot provide such subcellular specificity in their regulatory effects. In contrast, ribosomes, the likely site of miRNA action, distribute to various subcellular compartments, including the synapse, thereby providing potential subcellular resolution for miRNA-mediated regulatory control. miRNAs have indeed been recently implicated in synapse-specific functions (*26*, *27*). Taken together, speed, reversibility, and compartmentalization of miRNA-mediated control mechanisms predestine miRNAs to be involved in rapid, adaptive changes in gene expression to maintain homeostasis and to respond to specific environmental, nutrient, or neuronal signals.

In conclusion, TFs and miRNAs share many similarities and are broadly used in many different contexts. However, specific features of miRNA function may bias their use into more specialized regulatory niches, a notion that requires confirmation through detailed and comprehensive analyses of miRNA gene knockouts in the future.

References and Notes

1. O. Hobert, *Trends Biochem. Sci.* **29**, 462 (2004).
2. M. Plasterk, A. Gann, *Genes and Signals* (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, 2002).
3. A. Grimson *et al.*, *Mol. Cell* **27**, 91 (2007).
4. P. Sætrom *et al.*, *Nucleic Acids Res.* **35**, 2333 (2007).
5. M. J. Buck, J. D. Jeb, *Nat. Genet.* **38**, 1446 (2006).
6. M. Keddie *et al.*, *Cell* **131**, 1273 (2007).
7. H. Robins, Y. Li, R. W. Padgett, *Proc. Natl. Acad. Sci. U.S.A.* **102**, 4006 (2005).
8. S. Vasudevan, Y. Tong, J. A. Steitz, *Science* **318**, 1931 (2007).
9. R. S. Pillai, S. N. Bhattacharyya, W. Filipowicz, *Trends Cell Biol.* **17**, 118 (2007).
10. S. Gray, M. Levine, *Curr. Opin. Cell Biol.* **8**, 358 (1996).
11. E. H. Davidson, *Genomic Regulatory Systems* (Academic Press, San Diego, CA, 2001).
12. U. Alon, *An Introduction to Systems Biology: Design Principles of Biological Circuits* (Chapman & Hall/CRC, Boca Raton, FL, 2006).
13. R. J. Johnston Jr., S. Chang, J. F. Etchberger, C. O. Ortiz, O. Hobert, *Proc. Natl. Acad. Sci. U.S.A.* **102**, 12449 (2005).
14. J. Tsang, J. Zhu, A. van Oudenaarden, *Mol. Cell* **26**, 753 (2007).
15. J. M. Thomson *et al.*, *Genes Dev.* **20**, 2202 (2006).
16. E. J. Lee *et al.*, *RNA* **14**, 35 (2007).
17. Y. Kawahara *et al.*, *Science* **315**, 1137 (2007).
18. R. W. Carthew, *Curr. Opin. Genet. Dev.* **16**, 203 (2006).
19. E. A. Miska *et al.*, *PLoS Genet* **3**, e215 (2007).
20. www.wormbase.org
21. R. J. Johnston, O. Hobert, *Nature* **426**, 845 (2003).
22. J. Lu *et al.*, *Nature* **435**, 834 (2005).
23. E. Wienholds *et al.*, *Science* **309**, 310 (2005).
24. G. A. Wray *et al.*, *Mol. Biol. Evol.* **20**, 1377 (2003).
25. S. N. Bhattacharyya, R. Habermacher, J. Martine, E. I. Closs, W. Filipowicz, *Cell* **125**, 1111 (2006).
26. S. I. Ashraf, A. L. McLoon, S. M. Sclarsic, S. Kunes, *Cell* **124**, 191 (2006).
27. G. M. Schratt *et al.*, *Nature* **439**, 283 (2006).
28. K. C. Martin, M. Barad, E. R. Kandel, *Curr. Opin. Neurobiol.* **10**, 587 (2000).

29. I thank R. Mann, C. Desplan, N. Rajewsky, and members of my laboratory for comments on the manuscript. I regret that space limitations prevented a more extensive coverage of the literature. Work in my laboratory is funded by the NIH and the Howard Hughes Medical Institute.

Supporting Online Material

www.sciencemag.org/cgi/content/full/319/5871/1785/DC1
Figs. S1 and S2

10.1126/science.1151651

PERSPECTIVE

The Eukaryotic Genome as an RNA Machine

Paulo P. Amaral, Marcel E. Dinger, Tim R. Mercer, John S. Mattick*

The past few years have revealed that the genomes of all studied eukaryotes are almost entirely transcribed, generating an enormous number of non–protein-coding RNAs (ncRNAs). In parallel, it is increasingly evident that many of these RNAs have regulatory functions. Here we highlight recent advances that illustrate the diversity of ncRNA control of genome dynamics, cell biology, and developmental programming.

RNAs are an integral component of chromosomes and contribute to their structural organization (*1*, *2*). It is now becoming apparent that chromatin architecture and epigenetic memory are regulated by RNA-directed processes that, although the exact mechanisms are yet to be understood, involve the recruitment of histone-modifying complexes and DNA methyltransferases to specific loci (*3*). Whereas long non protein-coding RNAs (ncRNAs) have been classically implicated in the regulation of dosage compensation and genomic imprinting in animals (*4*), they seem to play a much broader role in the epigenetic control of developmental trajectories (*5*). For example, it was recently shown that 231 long ncRNAs associated with human *HOX* gene clusters are colinearly expressed along developmental axes (*5*), one of which, termed *HOTAIR*, transcribed from the *HOXC* locus, was studied in detail and found to recruit Polycomb complexes to repress gene expression in trans at the *HOXD* cluster (*5*) (Fig. 1). Other ncRNAs will likely perform similar functions, such as the intergenic transcripts from globin and antigen receptor loci, which have been associated with complex epigenetic phenomena (*6*, *7*).

Small ncRNAs have been consistently linked with heterochromatin formation via the RNA interference (RNAi) pathway (*8*), including Piwi-interacting RNAs (piRNAs) (*9*), which guide PIWI family proteins to control transposon activity from flies to vertebrates (*10*). However, piRNAs might also regulate euchromatin formation, given that PIWI is required for establishing euchromatin in certain subtelomeric regions in *Drosophila* (*11*).

Higher-level nuclear organization and chromosome dynamics are also regulated by ncRNAs in a variety of systems. For example, the formation of the kinetochore and centromeric heterochromatin in fission yeast is dependent on cell cycle–regulated centromeric repeat–derived RNAs and the RNAi pathway, whereas kinetochore assembly and chromosome segregation require the ribonuclease activity of a component of the exosome (*12*–*15*). These findings reveal an RNA-based mechanistic link between these processes in mitosis. In *Tetrahymena*, RNAs direct heterochromatin formation and DNA elimination via RNAi-dependent recruitment of Polycomb complexes and histone methylation (*16*). The RNAi pathway along with directed histone modifications also regulates the organization of the nucleolus in *Drosophila* (*17*).

Likewise, long ncRNAs direct programmed whole-genome rearrangements during ciliate differentiation (*18*). In mammals, transcription of long ncRNAs contributes to various processes including T cell receptor recombination (*7*), maintenance of telomeres (*19*, *20*), X-chromosome pairing required for dosage compensation (*21*) and inactive X-chromosome perinucleolar localization (*22*).

The functional organization of chromatin can also be regulated by ncRNAs derived from repetitive elements. In mice, bidirectional transcription of a retrotransposed SINE B2 sequence by RNA polymerase (RNAP)II and RNAPIII relocates the associated growth hormone locus into nuclear compartments and locally defines the heterochromatin-euchromatin boundary, regulating the expression of the gene during organogenesis (*23*) (Fig. 1). Given the abundance of transcribed repetitive sequences, this may represent a genome-wide strategy for the control of chromatin domains that may be conserved throughout eukaryotes (*23*–*25*). Moreover, such observations and others suggest that a large portion of the genome may, in fact, be functionally active and that transposon-derived sequences may not be reliable indices of the rate of neutral evolution (*26*).

Transcription

Noncoding RNAs can regulate transcription by interacting with transcription factors, RNAP, or

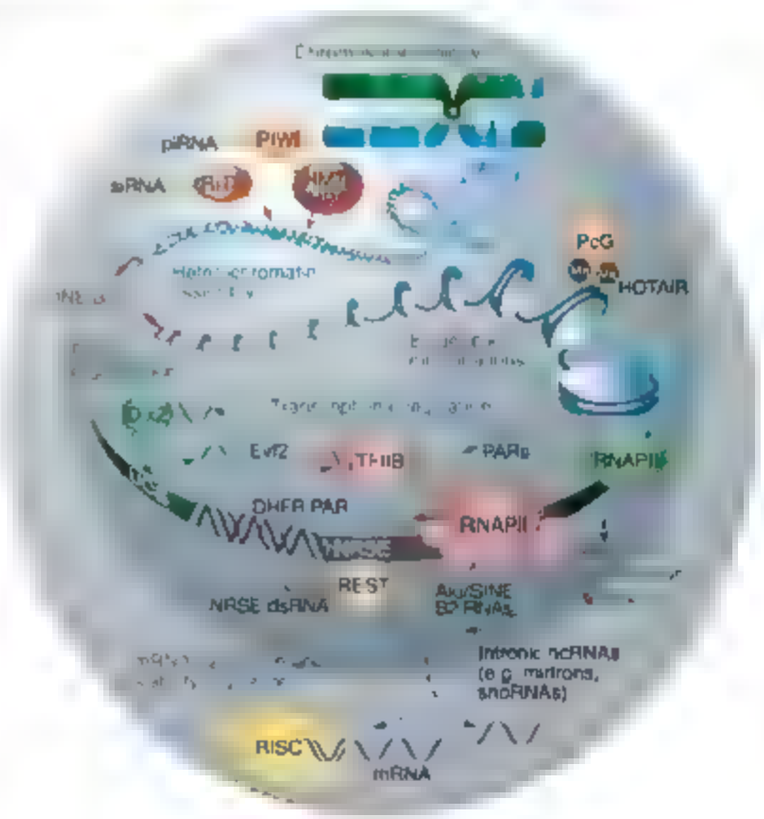


Fig. 1. Recent examples of the various levels of regulation of eukaryotic gene expression and cell biology by ncRNAs. dsRNA, double-stranded RNA; HMT, histone methyltransferases; HP1, heterochromatin protein 1; PARs, promoter-associated RNAs; PcG, Polycomb group proteins; RISC, RNA-induced silencing complex; RITS, RNA-induced initiation of transcriptional gene silencing; siRNA, small interfering RNA; TFIIB, transcription factor IIB; and UCE, ultraconserved element. See text for details, other acronyms, and references. For additional examples, see (*3*, *58*).

Institute for Molecular Bioscience, University of Queensland, St. Lucia QLD 4072, Australia.

*To whom correspondence should be addressed. E-mail: j.mattick@imb.uq.edu.au

DNA itself (Fig. 1). The small double-stranded RNA *NRSE* directs the activation of neuronal genes containing the upstream conserved *NRSE* sequence by triggering the REST transcriptional machinery (*27*), components of which are also targets of microRNAs (miRNAs) (*28*). Analogously, the ncRNA *Evf-2* is transcribed from an ultraconserved enhancer associated with the *Dlx-5/6* locus and interacts with *Dlx-2* transcription factor to induce enhancer activity (*29*). Indeed, it appears that most ultraconserved elements in the human genome (sequences that, in many cases, act as enhancers and have remained essentially unchanged throughout amniote evolution) (*30*) are transcribed in a regulated manner whose aberrant expression may be involved in pathological processes such as cancer (*31*).

In addition, promoter-directed sequence-specific RNAs, including miRNAs (*32*), have been shown to induce (*32*, *33*) or repress transcription (*34*, *35*), the latter involving epigenetic modification and targeting of promoter-associated low-copy RNA (*34*, *35*) [in one case by triplex formation (*35*), which appears common in human chromatin (*36*)]. The abundance of promoter-associated transcripts in the human (*37*, *38*), *Arabidopsis* (*39*), and yeast (*40*) genomes suggests that such examples may represent widespread phenomena.

RNAPIII can also produce various types of regulatory RNAs. In mice, heat shock induces RNAPIII transcription of a B2 SINE element producing a structured modular RNA that represses RNAPII in trans at specific loci (*41*). These same properties and function in human cells under heat shock were identified for ncRNAs derived from Alu elements, primate-specific SINEs that comprise 10.5% of our genome, and show a striking convergent evolution of SINE function as trans-repressors of gene expression (*42*).

RNA Processing and Translation

MiRNAs regulate a wide range of processes in animals and plants by directing translational repression or degradation of mRNAs (Fig. 1), with consequent effects on global regulatory circuits (*28*). However, miRNAs can also promote activation of gene expression under conditions such as stress, depending on the association with regulatory factors (*43*), and even activate translation during cell cycle arrest (*44*).

The formation and target specificity of miRNAs can be modulated by tissue-specific A-I editing (*45*) and RNA binding proteins (*46*, *47*). Furthermore, miRNAs may be differentially processed from both sense and antisense strands of the same hairpin or from sense and antisense transcripts from the same locus (*48*, *49*), which expands the potential of a single genomic locus to generate multiple miRNAs with different targets.

In addition, many miRNAs derive from introns of protein-coding genes, in some cases by splicing rather than the canonical Drosha pathway (*50*–*52*), thus, along with small nucleolar RNAs (snoRNAs), making up intronic parallel outputs of gene expression (*3*), which may be much more extensive than expected.

Other classes of ncRNAs can also act post-transcriptionally. In a recent report, a number of RNAPIII-transcribed short ncRNAs were found to have sequence complementarity to protein-coding genes and provided evidence for the existence of a sense-antisense–based regulatory network wherein RNAPIII transcripts control their RNAPII counterparts (Fig. 1) (*53*).

A World of Noncoding RNAs

The examples above provide proof-of-principle that RNA can regulate gene expression at many levels and by using a wide array of mechanisms (Fig. 1). The ENCODE project showed that at least 93% of analyzed human genome nucleotides are transcribed in different cells (*54*), with similar findings in mouse (*55*) and other eukaryotes, which indicate that there may be a vast reservoir of biologically meaningful RNAs that could greatly exceed the ~1.2% encoding proteins. A fraction of RNAs with short open reading frames (ORFs) potentially encodes peptides (*56*) but on the other side of the ledger many currently annotated ORFs are not conserved and may be false, which could reduce the number of protein-coding genes in the human genome (*57*).

There has been debate about whether these ncRNAs are (in the main) functional or simply noise. In some cases, it may be the transcript or merely the act of transcription, or both (*35*), that are relevant. Nevertheless, many observations indicate that substantial numbers of ncRNAs are intrinsically functional. These include the fact that many loci produce spliced (and alternatively spliced) transcripts that are developmentally regulated (*5*, *55*, *58*) and that at least some antisense and intergenic ncRNAs can function in trans (*5*, *41*, *53*, *59*). A large fraction of ncRNAs are expressed in specific regions of the brain, exhibiting precise cellular locations (*60*). Some mark new domains within the cell (*61*, *62*), which means that ncRNAs are also set to have a major impact in cell biology.

Comparative analyses indicate that ncRNA promoters are, on average, more conserved than those of protein-coding genes (*55*) and that ncRNA sequences, secondary structures, and splice site motifs have been subject to purifying selection (*63*, *64*). Moreover, many ncRNAs are evolving quickly (*65*), and some have undergone recent positive selection, as exemplified by *HAR1* RNA expressed in the human brain, which contains the sequence conserved in mammals that most rapidly diverged after the human-chimpanzee separation (*66*).

Given the functional versatility of RNAs, it is plausible that ncRNAs have represented a rich substrate for evolutionary innovations in eukaryotes. In support of this idea, regulatory RNAs are centrally involved in the ontogeny of many organisms, from unique developmental pathways in protozoa (*18*) to the control of conserved or clade-specific developmental regulators in multicellular animals (*5*, *52*). In addition, there is mounting evidence that RNA can transmit information intergenerationally, so as to mediate non-Mendelian inheritance of epigenetic changes in mice (*67*) and plants (*68*).

Although the need for large-scale approaches to explore the function of ncRNAs is evident, a glance at the genome browser will show non-coding expressed sequence tags associated with most genes of interest that may have regulatory functions. As an example, a recent study of ncRNAs associated with tumor suppressor genes focused on an RNA antisense to the *p15* gene and showed that it acts to alter histone methylation to silence expression of this gene, with important implications for cell differentiation and tumorigenesis (*59*).

Indeed, ncRNAs are already being identified as markers for cancer (*69*, *70*) and associated with other complex diseases such as coronary disease and diabetes (*71*, *72*). The elucidation of their function may significantly contribute to the understanding and treatment of such conditions. It may also transform our understanding of the genetic programming of multicellular organisms, particularly as it appears that regulation dominates the information content of complex systems (*3*).

References and Notes

1. J. A. Nickerson, G. Krochmalnic, K. M. Wan, S. Penman, *Proc. Natl. Acad. Sci. U.S.A.* **86**, 177 (1989).
2. A. Rodriguez-Campos, F. Azorin, *PLoS ONE* **2**, e1182 (2007).
3. J. S. Mattick, *J. Exp. Biol.* **210**, 1526 (2007).
4. P. K. Yang, M. I. Kuroda, *Cell* **128**, 777 (2007).
5. J. L. Rinn *et al.*, *Cell* **129**, 1311 (2007).
6. H. L. Ashe, J. Monks, M. Wijgerde, P. Fraser, N. J. Proudfoot, *Genes Dev.* **11**, 2494 (1997).
7. L. Alberteggui, M. S. Kiangel, *EMBO J.* **26**, 4383 (2007).
8. M. Buhler, O. Mouzed, *Nat. Struct. Mol. Biol.* **14**, 1041 (2007).
9. B. Brower-Toland *et al.*, *Genes Dev.* **21**, 2300 (2007).
10. A. A. Aravin, G. J. Hannon, J. Brennecke, *Science* **318**, 761 (2007).
11. H. Yin, H. Lin, *Nature* **450**, 304 (2007).
12. M. Buhler, W. Haas, S. P. Gygi, D. Mouzed, *Cell* **129**, 707 (2007).
13. H. Murakami *et al.*, *PLoS ONE* **2**, e317 (2007).
14. H. D. Folco, A. J. Pidoux, T. Urano, R. C. Allshire, *Science* **319**, 94 (2008).
15. E. S. Chen *et al.*, *Nature* **451**, 734 (2008).
16. Y. Liu *et al.*, *Genes Dev.* **21**, 1530 (2007).
17. J. C. Peng, G. H. Karpen, *Nat. Cell Biol.* **9**, 25 (2007).
18. M. Nowacki *et al.*, *Nature* **451**, 153 (2008).
19. S. Schoeftner, M. A. Blasco, *Nat. Cell Biol.* **10**, 228 (2008).
20. C. M. Azzalin, P. Reichenbach, L. Khoriauli, E. Giulotto, J. Lingner, *Science* **318**, 798 (2007).
21. N. Xu, M. E. Donohoe, S. S. Silva, J. T. Lee, *Nat. Genet.* **39**, 1390 (2007).
22. L. F. Zhang, K. D. Huynh, J. T. Lee, *Cell* **129**, 693 (2007).
23. V. V. Lunyak *et al.*, *Science* **317**, 248 (2007).
24. K. Nakaya, H. P. Cam, R. J. Maraia, S. I. Grewal, *Cell* **125** [illegible]
25. D. A. Willoughby, A. Visalli, R. G. Oshima, *J. Biol. Chem.* **275**, 759 (2000).
26. M. Pheasant, J. S. Mattick, *Genome Res.* **17**, 1245 (2007).
27. T. Kuwabara, J. Hsieh, K. Nakashima, K. Taira, F. H. Gage, *Cell* **116**, 779 (2004).

28. E. V. Makeyev, T. Maniatis, *Science* **319**, 1789 (2008).
29. J. Feng *et al.*, *Genes Dev.* **20**, 1470 (2006).
30. S. Stephen, M. Pheasant, I. V. Makunin, J. S. Mattick, *Mol. Biol. Evol.* **25**, 402 (2008).
31. G. A. Calin *et al.*, *Cancer Cell* **12**, 215 (2007).
32. R. F. Place, L. C. Li, D. Pookot, E. J. Noonan, R. Dahiya, *Proc. Natl. Acad. Sci. U.S.A.* **105**, 1608 (2008).
33. L. C. Li *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **103**, 17337 (2006).
34. J. Han, D. Kim, K. V. Morris, *Proc. Natl. Acad. Sci. U.S.A.* **104**, 12422 (2007).
35. I. Martianov, A. Ramadass, A. Serra Barros, N. Chow, A. Akoulitchev, *Nature* **445**, 666 (2007).
36. M. Ohta, T. Fukagawa, J. S. Lee, T. Ikemura, *Chromosoma* **111**, 201 (2002).
37. P. Kapranov *et al.*, *Science* **316**, 1484 (2007).
38. M. G. Guenther, S. S. Levine, L. A. Boyer, R. Jaenisch, R. A. Young, *Cell* **130**, 77 (2007).
39. J. A. Chekanova *et al.*, *Cell* **131**, 1340 (2007).
40. C. A. Davis, M. Ares Jr., *Proc. Natl. Acad. Sci. U.S.A.* **103**, 3262 (2006).
41. C. A. Espinosa, J. A. Goodrich, J. F. Kugel, *RNA* **13**, 583 (2007).
42. P. D. Mannor *et al.*, *Mol. Cell* **29**, 499 (2008).
43. A. K. Leung, P. A. Sharp, *Cell* **130**, 581 (2007).
44. S. Vasudevan, Y. Fong, J. A. Steitz, *Science* **318**, 1931 (2007).
45. Y. Kanahara *et al.*, *Science* **315**, 1137 (2007).
46. M. Keddie *et al.*, *Cell* **131**, 1273 (2007).
47. S. R. Viswanathan, G. Q. Daley, R. I. Gregory, *Science* 21 February 2008 (10.1126/science.1154040).
48. D. M. Tyler *et al.*, *Genes Dev.* **22**, 26 (2008).
49. A. Stark *et al.*, *Genome Res.* **17**, 1865 (2007).
50. J. G. Ruby, C. H. Jan, D. P. Bartel, *Nature* **448**, 83 (2007).
51. K. Okamura, J. W. Hagen, H. Duan, D. M. Tyler, E. C. Lai, *Cell* **130**, 89 (2007).
52. E. Bazzini, W. J. Chung, J. Willis, E. Cuppen, E. C. Lai, *Mol. Cell* **28**, 328 (2007).
53. A. Pagano *et al.*, *PLoS Genet.* **3**, e1 (2007).
54. E. Birney *et al.*, *Nature* **447**, 799 (2007).
55. P. Carninci *et al.*, *Science* **309**, 1559 (2005).
56. M. C. Frith *et al.*, *PLoS Genet.* **2**, e52 (2006).
57. M. Clamp *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **104**, 19428 (2007).
58. K. V. Prasanth, D. L. Spector, *Genes Dev.* **21**, 11 (2007).
59. W. Yu *et al.*, *Nature* **451**, 202 (2008).
60. T. R. Mercer, M. E. Dinger, S. M. Sunkin, M. F. Mehler, J. S. Mattick, *Proc. Natl. Acad. Sci. U.S.A.* **105**, 716 (2008).
61. M. Sone *et al.*, *J. Cell Sci.* **120**, 2498 (2007).
62. H. Royo *et al.*, *Mol. Biol. Cell* **18**, 2017 (2007).
63. E. Tomasson *et al.*, *Genome Res.* **18**, 242 (2008).
64. J. Ponjavic, C. P. Ponting, G. Lunter, *Genome Res.* **17**, 556 (2007).
65. K. C. Pang, M. C. Frith, J. S. Mattick, *Trends Genet.* **22**, 1 (2006).
66. K. S. Pollard *et al.*, *Nature* **443**, 167 (2006).
67. M. Rasoulzadegan *et al.*, *Nature* **441**, 469 (2006).
68. C. J. Hale, J. L. Stanley, S. M. Gross, J. B. Hollick, *PLoS Biol.* **5**, 2156 (2007).
69. R. Lin, S. Maeda, C. Liu, M. Karin, T. S. Edgington, *Oncogene* **26**, 851 (2007).
70. D. S. Perez *et al.*, *Hum. Mol. Genet.* **17**, 642 (2007).
71. N. Ishii *et al.*, *J. Hum. Genet.* **53**, 1087 (2008).
72. H. M. Broadbent *et al.*, *Hum. Mol. Genet.* **17**, 806 (2007).

73. The authors acknowledge the support of the Australian Research Council and the New Zealand Foundation for Research Science and Technology.

10.1126/science.1155472

PERSPECTIVE

Multilevel Regulation of Gene Expression by MicroRNAs

Eugene V. Makeyev and Tom Maniatis

MicroRNAs (miRNAs) are ~22-nucleotide-long noncoding RNAs that normally function by suppressing translation and destabilizing messenger RNAs bearing complementary target sequences. Some miRNAs are expressed in a cell- or tissue-specific manner and may contribute to the establishment and/or maintenance of cellular identity. Recent studies indicate that tissue-specific miRNAs may function at multiple hierarchical levels of gene regulatory networks, from targeting hundreds of effector genes incompatible with the differentiated state to controlling the levels of global regulators of transcription and alternative pre-mRNA splicing. This multilevel regulation may allow individual miRNAs to profoundly affect the gene expression program of differentiated cells.

There is a surprisingly small difference in the number of protein-encoding genes between organisms of vastly different morphological and behavioral complexity. A plausible explanation for this paradox may be in the increased elaboration of gene regulatory networks at the levels of transcription (*1*) and alternative pre-mRNA splicing (*2*). More recently, miRNAs have been proposed to play a role in the expansion of organismal complexity (*3*). Indeed, some miRNAs are expressed in a cell- or tissue-specific manner during embryonic development, suggestive of a role in cellular differentiation (*4*) [also see essay by O. Hobert in this issue (*5*)]. Certain of these tissue-specific miRNAs, such as miR-1 and miR-124, which are expressed in muscle cells and neurons, respectively, have been shown to stimulate differentiation of the corresponding cell types (*6*–*8*).

Department of Molecular and Cellular Biology, Harvard University, Cambridge, MA 02138, USA. E-mail: makeyev@mcb.harvard.edu (E.V.M.); maniatis@mcb.harvard.edu (T.M.)

Targeting Gene Batteries

Several distinct molecular mechanisms may underlie the biological functions of miRNAs. Individual miRNAs can repress large sets of mRNAs that are not required at a particular developmental stage (*9*, *10*). A number of these miRNA targets fall into the category of "gene batteries," sets of functionally related effector genes that represent outputs of gene regulatory networks (*11*). Indeed, when miR-124 is introduced into nonneuronal mammalian cells a preferential reduction in the amounts of multiple nonneuronal mRNAs, for example, those encoding proteins required for cell proliferation or neural stem cell function, is observed (*7*, *9*, *12*). Conversely, depletion of miR-124 from primary neurons leads to the accumulation of a number of nonneuronal mRNA targets (*13*). A similar mode of action has been reported for miR-1 (*9*). Thus, in cells undergoing differentiation, miRNAs can effectively deplete unwanted mRNAs left over from progenitor cells. Furthermore, miRNAs may regulate a fraction of mRNA targets at the level of translation without affecting their stability (*14*, *15*).

Targeting Regulators of Transcription

MiRNAs can also control the expression of critical transcriptional regulators, as first shown by the repression of transcription factor lin-14 by miRNA lin-4 (*16*, *17*). Another example is miR-124, which targets mRNA of the small C-terminal domain phosphatase 1 (SCP1/CTDSP1) protein, a component of the REI-silencing transcription repressor (REST/NRSF) (*6*) (Fig. 1). REST represses the transcription of a large number of neuron-specific genes in nonneural cells (*13*). Thus, by reducing SCP1 expression in differentiating neurons, miR-124 may facilitate the derepression of the neuronal transcription program (*6*). The REST complex may also repress miR-124 gene expression and thus establish a double-negative feedback loop between miR-124 and the REST complex (*6*, *13*) (Fig. 1). Similarly, miR-1 and another muscle-specific miRNA, miR-133, regulate several transcription factors essential for muscle development (*4*, *8*).

Targeting Regulators of Alternative Splicing

Recent studies have identified yet another activity of miRNAs: the induction of large-scale changes in gene expression by targeting global regulators of alternative pre-mRNA splicing (*7*, *18*). In early muscle cell precursors, a repressor of alternative splicing called polypyrimidine tract-binding protein 1 (PTBP1/PTB/hnRNP-I) and its homolog, PTBP2 (nPTB/brPTB/PTBLP), repress the inclusion of a number of muscle-specific exons into mature mRNAs. However, upon myotube differentiation, PTBP1 and PTBP2

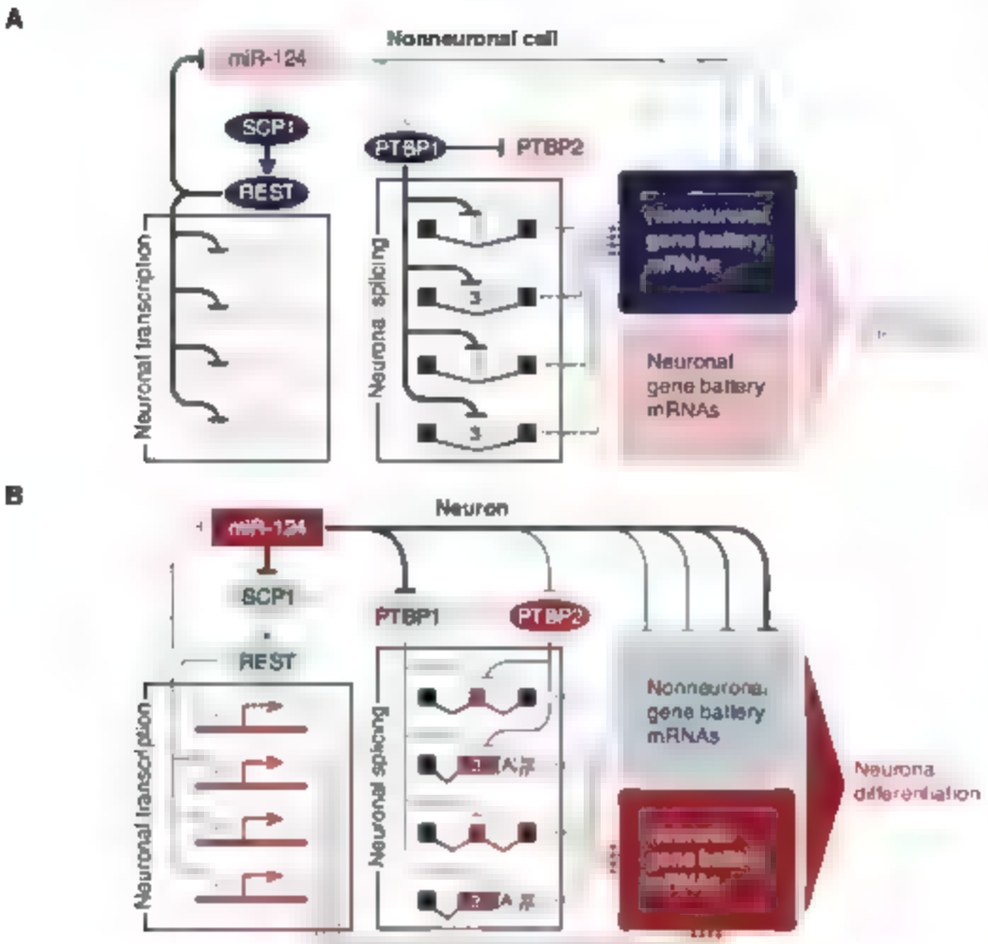


Fig. 1. MiR-124 controls an extensive gene regulatory network. Active regulatory interactions are shown as solid black lines. Inactive interactions are in gray. Weaker regulatory interactions are rendered as thinner lines. Gene battery expression inputs to the cellular transcriptome are indicated as dashed lines. Nonneuronal elements are colored in dark blue (present) or light blue (absent); neuron-specific elements are colored in corresponding shades of red. This is a simplified diagram that does not account for potential inputs from other neuron-specific miRNAs, as well as the recently identified miRNA function in translational activation (*15*). Furthermore, PTBP1 and PTBP2 may also activate subsets of alternative exons by using a yet-to-be understood mechanism (*19*). (**A**) In nonneuronal cells or neural precursors, miR-124 is either absent or present at a low amount, which allows efficient expression of global repressors of neuronal transcription and alternative pre-mRNA splicing, as well as nonneuronal gene batteries. (**B**) In differentiating neurons, miR-124 quantities increase, leading to the down-regulation of corresponding repressors, thus allowing the production of the corresponding neuronal proteins. Inversely, the nonneuronal battery genes are directly down-regulated by miR-124.

protein quantities are reduced, leading to increased inclusion of their target exons during splicing. This splicing switch is regulated, at least in part, by miRNAs: The aforementioned miR-133 strongly suppresses PTBP2 production, whereas miR-1 and its sequence homolog miR-206 may also contribute to down-regulation of the PTBP1 and PTBP2 quantities during muscle development (*18*).

The amounts of PTBP1 and PTBP2 are also regulated by miRNAs during the development of the nervous system (*7*, *19*). PTBP1 functions in this context as a repressor of neuron-specific alternative exon inclusion, and it is expressed in neural precursors as well as many other types of nonneuronal cells. However, in differentiating and mature neurons, PTBP1 quantities decrease, leading to the inclusion of a number of neuron-specific alternative exons in mature mRNA (*19*). Similar to the regulation in muscle cells, the reduced PTBP1 expression in neurons is mediated by miR-124, which interacts with conserved and nonconserved cognate target sites in the 3′ untranslated region (3′UTR) of the PTBP1 mRNA (*7*) (Fig. 1).

In addition to other alternative exons, PTBP1 represses the inclusion of exon 10 of PTBP2 pre-mRNA (*7*, *19*, *20*). When this exon is skipped (i.e., excluded from mature mRNA), PTBP2 mRNA acquires a premature termination codon and as a result is degraded by the nonsense-mediated decay machinery (*7*, *19*, *20*). During neuronal differentiation, miR-124 quantities increase, thus reducing PTBP1 amounts and resulting in the accumulation of correctly spliced PTBP2 mRNA, which in turn leads to a substantial increase in PTBP2 protein. Notably, although PTBP1 and PTBP2 are closely related homologs, PTBP1 appears to be a much stronger repressor of neuron-specific alternative exons (*7*, *19*). Thus, the reduction in PTBP1 and increase in PTBP2 results in a global switch from nonneuronal to neuron-specific alternative splicing patterns, leading to the production of neuron-specific protein isoforms. Interestingly, PTBP2 mRNA contains conserved miR-124 binding sites that allow miR-124 to suppress PTBP2 expression, albeit less efficiently than PTBP1 (*7*). This regulation may provide a mechanism that dampens PTBP2 expression (Fig. 1).

The increasing diversity of cellular differentiation in metazoans, accompanied by an increase in the complexity of gene regulatory networks, must have required a mechanism to prevent interference between spatially or temporally adjacent gene expression programs. The above examples argue that at least some miRNAs play an important role in this mechanism by effectively rewiring the cell-specific networks at all levels of the regulatory hierarchy, from the gene battery to global regulators of transcription and alternative splicing. It seems likely that other examples of this multilevel regulation will be revealed as target repertoires of other miRNAs are determined.

References and Notes

1. M. Levine, R. Tjian, *Nature* **424**, 147 (2003).
2. T. Maniatis, B. Tasic, *Nature* **418**, 236 (2002).
3. R. Nina, F. J. Slack, *Curr. Opin. Genet. Dev.* **17**, 145 (2007).
4. Y. Zhao, D. Srivastava, *Trends Biochem. Sci.* **32**, 189 (2007).
5. O. Hobert, *Science* **319**, 1785 (2008).
6. J. Visvanathan, S. Lee, B. Lee, J. W. Lee, S. K. Lee, *Genes Dev.* **21**, 744 (2007).
7. E. V. Makeyev, J. Zhang, M. A. Carrasco, T. Maniatis, *Mol. Cell* **27**, 435 (2007).
8. J. F. Chen *et al.*, *Nat. Genet.* **38**, 228 (2006).
9. L. P. Lim *et al.*, *Nature* **433**, 769 (2005).
10. A. J. Giraldez *et al.*, *Science* **312**, 75 (2006); published online 15 February 2006 (10.1126/science.1122689).
11. E. H. Davidson, *The Regulatory Genome: Gene Regulatory Networks in Development and Evolution* (Academic Press, San Diego, CA, 2006).
12. X. Cao, S. Y. Pfaff, F. H. Gage, *Genes Dev.* **21**, 531 (2007).
13. C. Conaco, S. Otto, J. J. Han, G. Mandel, *Proc. Natl. Acad. Sci. U.S.A.* **103**, 2422 (2006).
14. F. V. Karginov *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **104**, 1929 (2007).
15. S. Vasudevan, Y. Tong, J. A. Steitz, *Science* **318**, 1931 (2007); published online 29 November 2007 (10.1126/science.1149460).
16. R. C. Lee, R. L. Feinbaum, V. Ambros, *Cell* **75**, 843 (1993).
17. B. Wightman, I. Ha, G. Ruvkun, *Cell* **75**, 855 (1993).
18. P. L. Boutz, G. Chawla, P. Stoilov, D. L. Black, *Genes Dev.* **21**, 71 (2007).
19. P. L. Boutz *et al.*, *Genes Dev.* **21**, 1636 (2007).
20. R. Spellman, M. Llorian, C. W. Smith, *Mol. Cell* **27**, 420 (2007).
21. We thank S. Ollikerenko, S. Buchanan, B. Friedman, E. Morimoto, and B. McCallum for critical comments. This effort was supported by NIH (grant 2R01NS043915-27 TM.) and Leukemia and Lymphoma Society (E.V.M.).

10.1126/science.1152326

PERSPECTIVE

Transcription Regulation Through Promoter-Proximal Pausing of RNA Polymerase II

Leighton J. Core and John T. Lis*

Recent work has shown that the RNA polymerase II enzyme pauses at a promoter-proximal site of many genes in *Drosophila* and mammals. This rate-limiting step occurs after recruitment and initiation of RNA polymerase II at a gene promoter. This stage in early elongation appears to be an important and broadly used target of gene regulation.

Most eukaryotic genes that encode messenger RNAs are subject to primary regulation at the level of transcription. Transcription of these RNAs consists of a series of distinct phases during which RNA polymerase II (Pol II) is recruited to a gene promoter, initiates transcription, escapes from the promoter and any proximal pause sites, elongates the RNA transcript, and eventually terminates transcription (*1*). These phases can each be further dissected into discrete biochemical steps that are potential targets of regulation. To understand how regulation works for a particular gene, it is important to identify which of these steps is rate-limiting and how signal-responsive activators and repressors act on them mechanistically.

Mechanistic studies of transcription regulation in eukaryotes have predominantly focused on the stages of preinitiation complex formation and the initiation of transcription; however, examples of regulation during elongation have been noted repeatedly over the past 25 years. A compelling early study by Chambon and colleagues showed that transcriptionally engaged Pol II at the 5′ end of the adult β-globin genes persisted after the genes were supposedly shut down in mature erythrocytes (*2*). A focused series of studies identified Pol II that was transcriptionally engaged but paused (*3*) 20 to 50 bases downstream of the transcription start site of uninduced *Drosophila Hsp70* (*4*) and several other *Drosophila* and mammalian genes (*5*, *6*). This promoter-proximal pausing, which has similarities to a transcription regulatory mechanism observed in prokaryotes (*7*), was postulated to be a rate-limiting step in gene regulation (*4*, *8*). Whereas promoter-proximal pausing was being established at a modest number of model genes in metazoans (*5*, *6*), regulation at an earlier stage—recruitment of Pol II to the promoter—was proving to be the general rule in the yeast *Saccharomyces cerevisiae* (*9*, *10*). Thus, observations where Pol II was transcriptionally engaged at unactivated promoters seemed unusual and were generally viewed as exceptions.

Department of Molecular Biology and Genetics, Cornell University, Ithaca NY 14853, USA.

*To whom correspondence should be addressed. E-mail: jtl10@cornell.edu

Generality of Promoter-Proximal Pausing

Although Pol II levels at a gene promoter generally correlate with mRNA levels in *S. cerevisiae* (*11*), several recent genome-wide analyses have revealed that this is not always the case in mammalian and *Drosophila* cells (*12*–*15*). These studies used the chromatin immunoprecipitation assay coupled with genomic microarray technologies ("ChIP-chip") to examine Pol II density along genes. These studies found that about 20 to 30% of genes have enriched Pol II density at the 5′ end relative to the body of the genes. This class included genes with either detectable or undetectable expression. Identification of this latter subclass, which has Pol II bound without full-length transcript production, suggests that a postrecruitment step of the transcription cycle is rate-limiting at these genes. Whereas the ChIP assay can detect the density of Pol II across a gene, it cannot necessarily determine whether Pol II is transcriptionally engaged; that is, the 5′-skewed distribution of Pol II could represent Pol II in either the preinitiation form or the initiated-but-paused form.

Three of the genome-wide studies presented additional assays of permanganate footprinting, which maps the transcription bubble associated with a transcriptionally engaged Pol II, or analysis of short RNA products as evidence that Pol II had progressed beyond initiation at multiple candidate genes (*13*–*15*). Although the validated genes in these studies were mainly in the class of low or nondetectable expression levels, it is important to emphasize that highly expressed genes were also

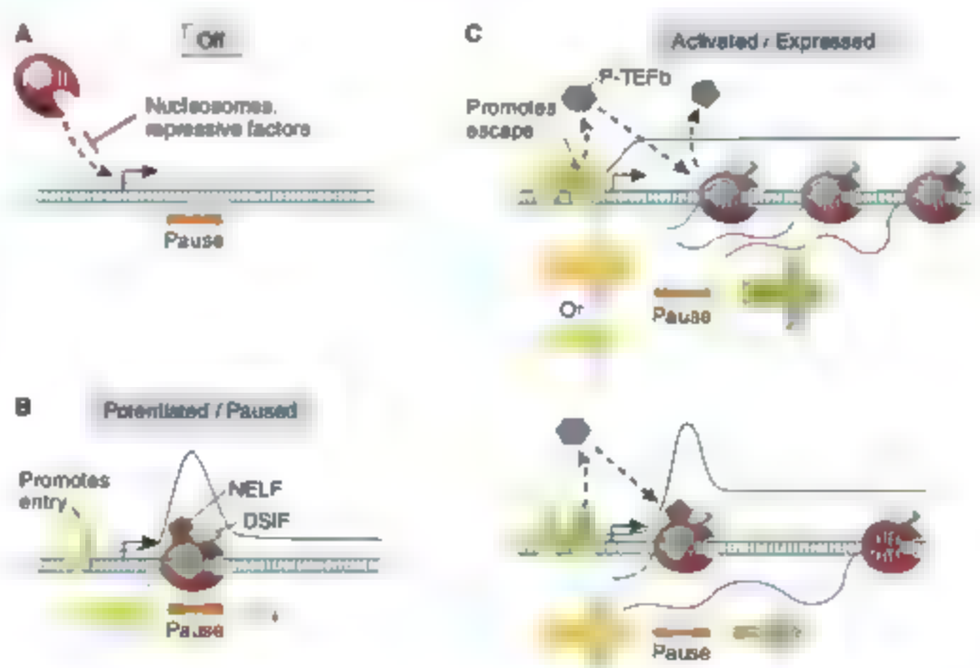


Fig. 1. Regulation of entry and escape of Pol II at pause sites. The rate of pause site entry (yellow arrows—wide arrow, fast entry; narrow arrow, slow entry) is defined as the rate at which Pol II (red) enters a pause site when it is freely accessible. The relative rates of entry and escape (green arrow) produce the observed patterns of Pol II density (blue line). (**A**) Pol II cannot access the promoter and transcription is "off." (**B**) A potentiated state through the setup of a promoter-proximal paused Pol II by factors that promote entry (yellow oval). NELF (pentagon) and DSIF (blue oval) stabilize the paused Pol II. (**C**) Fully activated transcription requires factors that promote escape (green rectangle). Also, single factors can have one or both types of activation domains that in turn can be regulated by reversible modifications and associations.

identified as candidates for pausing. Thus, regulation at the level of pausing appears to occur broadly and over a large dynamic range of transcript production.

Pausing Is a Target of Regulation

Several in vitro studies have shown that promoter-proximal pausing is a natural process that Pol II undergoes even in the absence of auxiliary factors. The DNA template and nascent RNA sequences are proposed to affect a position-dependent structural change in the transcription complex during early elongation (*16*, *17*). Such a conformational change may be necessary for achieving the fully processive form capable of transcribing long distances without disengaging the template or nascent RNA. The extent of intrinsic pausing in vivo is unclear, but the position relative to the start site is coincident with the action of the known pausing factors DSIF (DRB sensitivity inducing factor) and NELF (negative elongation factor) (*18*), which appear to further stabilize Pol II in the paused form. These factors have been the subject of recent reviews (*1*, *6*, *19*).

Entry of Pol II into a promoter-proximal pause site requires that the transcription machinery must first gain access to the promoter and initiate transcription. Escape from the pause site occurs when Pol II moves into productive elongation, which clears Pol II from the promoter and allows sufficient space for another Pol II complex to initiate transcription. The relative rates of entry and escape combine to determine the effective level of pausing at a gene (Fig. 1). High entry levels in combination with low escape rates result in increased dwell time at promoter-proximal pause sites, this is reflected by a high density of Pol II at the 5′ end relative to downstream regions of genes as seen by ChIP analysis (Fig. 1, B and C, bottom). Low Pol II dwell times are observed when the entry rate is less than or equal to pause site escape, the result is a more uniform occupancy of Pol II over the gene (Fig. 1C, top).

The differential rate of Pol II entry and escape at a pause site is well documented for the *Hsp70* gene of *Drosophila*. At this gene, GAGA factor is required for efficient pause site entry under nonactivating conditions, whereas binding of activated heat shock factor (HSF) is required for stimulation of escape and full activation of the gene (*6*). Similar, although generally less defined, examples of cooperating activators that stimulate different rate-limiting steps have been found in mammalian and *Drosophila* systems (*5*, *20*–*22*). A major implication that arises from the recent genome-wide studies is that activators that stimulate pause site entry without escape may function to potentiate transcription at a large number of promoters for later activation by additional factors. Thus theoretically imparts combinatorial control over transcription output, allowing cells to integrate more diverse signaling pathways and to synergistically up-regulate genes rapidly when needed.

Activators that stimulate escape do so by recruiting factors that directly modulate the transcription complex, and/or possibly by manipulating the chromatin environment such that transcription through nucleosomes is possible. The primary executor of escape from pausing is the kinase activity of positive transcription elongation factor b (P-TEFb) (*19*, *23*). This factor phosphorylates multiple targets within the transcription complex, including Pol II, NELF, and DSIF, and is crucial for relief of the NELF- and DSIF-dependent block to transcription elongation (Fig. 1). At this transition, NELF dissociates from the transcription complex, but the modified DSIF remains associated and enhances elongation. Not surprisingly, cells have developed a number of ways to bring P-TEFb to genes, through direct interaction with activators or through interactions with proteins that are brought in during activation (*6*, *19*).

Pausing and the Connection to RNA Processing

Promoter-proximal pausing occurs at a point where it may serve to coordinate transcription elongation with pre-mRNA processing (*1*). Indeed, pausing is coincident with mRNA capping (*6*), which stabilizes the RNA and is important for downstream processing events (*1*). Additionally, Pol II undergoes stepwise phosphorylation of the C-terminal domain (CTD) of its largest subunit as it progresses from its entry mode to its elongation mode, and this brings about the accompanying changes in its entourage of associated pre-mRNA processing proteins (*1*).

Not Just Repression: Potentiation of Transcription

Considerable evidence suggests that maintenance of pausing at a promoter is key for full activation of a gene. Studies on the *Drosophila Hsp70* and human *FOS* and *MYC* genes have shown that removal of the sequences that cause pausing result in decreased transcription factor accessibility and defective activation (*24*–*26*). How a paused Pol II grants accessibility of promoter DNA to regulatory factors is unclear, but it is possible that Pol II exerts this effect either by directly preventing nucleosomes from obstructing DNA binding sites, or by recruiting other factors that modify the chromatin architecture around the promoter. Of the genes identified by the genomic studies as likely having a paused polymerase in *Drosophila*, genes that respond rapidly to developmental and cell signaling were overrepresented (*14*, *15*). Thus raises the likelihood that potentiation through pausing before activation is a fundamental step for rapidly controlling developmental programs, as suggested by recent studies (*15*, *27*).

Transcription regulation is a multistep process that is controlled at the level of recruitment, initiation, pausing, and elongation of RNA polymerase II. A number of genome-scale studies have identified large classes of genes that are likely to be regulated by promoter-proximal pausing, and thus have provided us with a large set of model genes with which to study distinct aspects of this mode of regulation. Future investigations should focus on determining how promoter-specific binding proteins affect the transition between initiation and pausing, as well as the transition between pausing and productive elongation; the results will provide important insights into the role of cell signaling events in the mechanics of transcription regulation.

References and Notes

1. R. J. Sims 3rd, R. Belotserkovskaya, D. Reinberg, *Genes Dev.* **18**, 2437 (2004).
2. P. Gariglio, M. Bellard, P. Chambon, *Nucleic Acids Res.* **9**, 2589 (1981).
3. Although a large fraction of promoter-proximal Pol II can elongate in nuclear run-on assays, some Pol II may be backtracked and cannot readily elongate. Therefore the broader term of stalling, which includes arrest and pause, has also been used.
4. A. E. Rougvie, J. T. Lis, *Cell* **54**, 795 (1988).
5. D. L. Bentley, *Curr. Opin. Genet. Dev.* **5**, 210 (1995).
6. A. Saunders, L. J. Core, J. T. Lis, *Nat. Rev. Mol. Cell Biol.* **7**, 557 (2006).
7. J. W. Roberts *et al.*, *Cold Spring Harb. Symp. Quant. Biol.* **63**, 319 (1998).
8. J. Lis, *Cold Spring Harb. Symp. Quant. Biol.* **63**, 347 (1998).
9. M. Keaveney, K. Struhl, *Mol. Cell* **1**, 917 (1998).
10. M. Planiol, A. Gann, *Nature* **386**, 569 (1997).
11. F. Robert *et al.*, *Mol. Cell* **16**, 199 (2004).
12. T. H. Kim *et al.*, *Nature* **436**, 876 (2005).
13. M. G. Guenther, S. S. Levine, L. A. Boyer, R. Jaenisch, R. A. Young, *Cell* **130**, 77 (2007).
14. G. W. Muse *et al.*, *Nat. Genet.* **39**, 1507 (2007).
15. J. Zeitlinger *et al.*, *Nat. Genet.* **39**, 1512 (2007).
16. M. Pal, D. McKean, D. S. Luse, *Mol. Cell. Biol.* **21**, 5815 (2001).
17. A. Ujvari, M. Pal, D. S. Luse, *J. Biol. Chem.* **277**, 32527 (2002).
18. Y. Yamaguchi *et al.*, *Cell* **97**, 41 (1999).
19. B. M. Peterlin, D. H. Price, *Mol. Cell* **23**, 297 (2006).
20. J. Blau *et al.*, *Mol. Cell. Biol.* **16**, 2044 (1996).
21. T. Sawado, J. Halow, M. A. Bender, M. Groudine, *Genes Dev.* **17**, 1009 (2003).
22. Y. V. Wang, H. Tang, D. S. Gilmour, *Mol. Cell. Biol.* **25**, 3543 (2005).
23. N. F. Marshall, D. H. Price, *Mol. Cell. Biol.* **12**, 2078 (1992).
24. L. S. Shopland, K. Hirayoshi, M. Fernandes, J. T. Lis, *Genes Dev.* **9**, 2756 (1995).
25. A. Krumm, L. B. Hickey, M. Groudine, *Genes Dev.* **9**, 559 (1995).
26. J. Fhaz, M. C. Bassi, S. Pinaud, F. Mirkovitch, *Gene* **255**, 185 (2000).
27. X. Wang, C. Lee, D. S. Gilmour, J. P. Gergen, *Genes Dev.* **21**, 1031 (2007).

28. We thank members of the Lis lab and J. Roberts for comments on the manuscript. Recent reviews cited above provide access to the primary literature that unfortunately could not be cited in this concise Perspective. Supported by an NIH grant.

10.1126/science.1150843

PERSPECTIVE

Gene Regulation in the Third Dimension

Job Dekker*

Analysis of the spatial organization of chromosomes reveals complex three-dimensional networks of chromosomal interactions. These interactions affect gene expression at multiple levels, including long-range control by distant enhancers and repressors, coordinated expression of genes, and modification of epigenetic states. Major challenges now include deciphering the mechanisms by which loci come together and understanding the functional consequences of these often transient associations.

In compact genomes of organisms such as yeast, a gene and its regulatory elements form an uninterrupted genomic segment that constitutes a "regulatory expression unit." However, in more complex genomes, such as those of human and mouse, genes and their regulatory elements can be dispersed over many hundreds of kilobases (*1*, *2*). It has long been hypothesized that communication between widely spaced genomic elements can be facilitated by the spatial organization of chromosomes that brings genes and their regulatory elements in close proximity (Fig. 1A).

The organization of chromosomes has been studied by microscopy and more recently by chromosome conformation capture (3C) (*3*). 3C is a molecular technique that uses formaldehyde cross-linking and locus-specific PCR to detect physical contacts between genomic loci. These approaches are complementary, with microscopy providing information on single cells, but with relatively low resolution, and 3C allowing much higher resolution analyses, but requiring larger cell populations. 3C and microscopy studies confirm that long-range chromosomal interactions are widespread, which suggests a high level of communication between dispersed genomic elements.

Spatial Assembly of Expression Units

Well-characterized examples of spatial association of genomic elements involve interactions between enhancers and target genes. An example is that of the β-globin locus. The locus contains several β-globin–like genes that are regulated by a single cis-acting element, the locus control region (LCR), which is located about 10 to 60 kb upstream of the globin gene promoters. The LCR was found to physically associate with the active globin gene (*4*). Many more examples of long-range looping events have been identified, e.g., in the α-globin locus (*5*, *6*) and the interleukin gene cluster (*7*), and also in controlling single genes [e.g. (*8*, *9*)].

Highly specific associations between loci located on separate chromosomes have also been described. These trans-interactions can be between enhancers and putative target genes, as in the case of olfactory receptor genes (*10*). However, in other cases, they appear to play a role in a higher level of gene control to coordinately regulate multiple loci (Fig. 1B). One example is the association between the T helper 2 cytokine locus on mouse chromosome 11 and the interferon γ gene on chromosome 10 (*11*). Expression of these loci is mutually exclusive and interaction between them may provide an opportunity to initiate or enforce opposite epigenetic states.

The process of mammalian X-chromosome inactivation involves a specific trans-association between the X chromosomes. Female cells carry two X chromosomes, one of which is mostly silenced so that expression levels of X-linked genes are comparable to those in male cells. X inactivation is initiated at the X-inactivation center. Recently, a transient interaction between the two X-inactivation centers was detected during the developmental stages at which X inactivation is initiated (*12*, *13*). Analysis of mutations in the inactivation centers showed that their association is intimately involved in the X-inactivation process. X chromosome pairing provides an elegant mechanism for counting the number of X chromosomes and for ensuring that their epigenetic fates are linked so that when one chromosome is inactivated the other is not.

These observations suggest an interesting model of what constitutes a "regulatory expression unit" in complex genomes. Whereas in compact genomes, genes and their regulatory elements cluster along the linear genome sequence, in more complex genomes, expression units can be assembled by spatial clustering of genes and distant regulatory elements (Fig. 1). This mode of de novo assembly of expression units could provide additional levels of gene regulation by allowing combinatorial association of genes and sets of regulatory elements. For example, for imprinted loci, maternal and paternal alleles associate with different elements to assemble into distinct expression units (*14*).

Chromosomal Interactions Are Transient

Many of the observations of long-range interactions have been made using 3C and its variations. Performing 3C is relatively simple, but it has proven more complicated to interpret

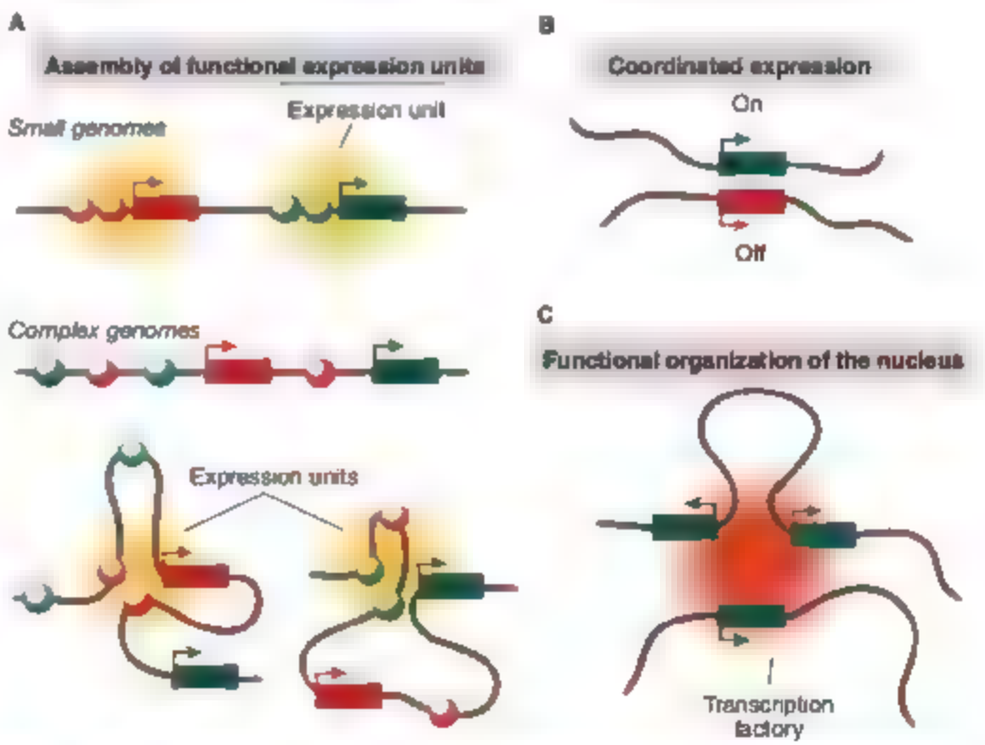


Fig. 1. Spatial assemblies. (**A**) Linearly defined expression units in compact genomes and spatially assembled expression units in complex genomes. (**B**) Association between coordinately expressed genes. (**C**) Colocalization of genes at subnuclear structures, such as transcription factories. Circles, regulatory elements; rectangles, genes. Arrows indicate direction of transcription.

Program in Gene Function and Expression and Department of Biochemistry and Molecular Pharmacology, University of Massachusetts Medical School, 364 Plantation Street, Worcester, MA 01605–0103, USA.

*To whom correspondence should be addressed. E-mail: Job.Dekker@umassmed.edu

the results, as has been discussed in several reviews (*15*, *16*). In particular, although many of the chromosomal interactions detected with 3C have been confirmed by microscopy, it is difficult to relate 3C signals to actual frequencies of association. In many cases, the frequency of colocalization is rather low (less than 10% of cells at a given point in time), in accordance with the fact that chromosome conformation is dynamic and highly variable among individual cells. Therefore, the common use of rather rigid looping models to describe these associations, although appealing, can be misleading because these models do not reflect the highly transient nature of long-range interactions.

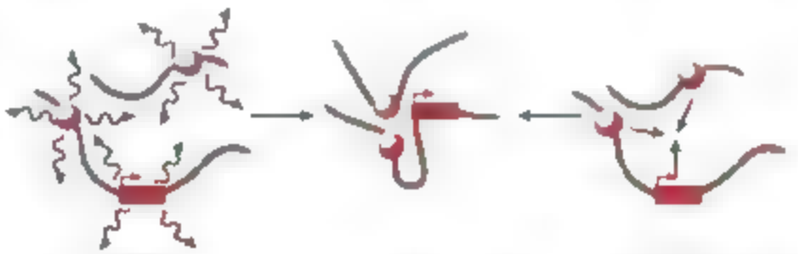


Fig. 2. Passive and active models for bringing loci together. Circles and rectangles represent regulatory elements and genes. Wavy arrows indicate random diffusion. Straight arrows indicate active and directed movement.

Functions of Chromosomal Interactions

Observing a specific association between two loci does not by itself reveal a function for that interaction. Additional approaches such as knock-down of proteins (e.g., transcription factors) that mediate the interaction or deletion of the regulatory element can reveal causal relations between long-range interactions and gene regulation. Another powerful approach is to analyze colocalization of loci by in situ hybridization combined with simultaneous visualization of RNA production at the gene to determine whether the interaction is correlated in time with gene transcription at the level of single cells. It should be noted that interactions have been observed that correlated with gene transcription but that deletion of the interacting regulatory element did not affect expression (*10*, *17*). Although this could indicate that the interaction is not relevant, it could equally reflect our very limited understanding of the role of chromosomal associations in genome regulation.

How do chromosomal associations affect gene expression? Enhancer-promoter interactions could aid in stable recruitment of components of the transcription machinery to the promoter. In addition, enhancer-bound enzymatic activities could be brought in contact with promoter complexes that are then modified, e.g., phosphorylated or methylated, which leads to modulation of promoter activity. Other types of interactions, such as those between the X-inactivation centers, could allow coordinated assembly of two distinct protein complexes on the interacting partners. Alternatively, given the very transient nature of these associations, the two loci may acquire distinct but stable marks, e.g., DNA methylation, that direct assembly of protein complexes at later time points when the loci no longer interact.

How Do Loci Get Together?

Several models have been proposed (*18*) by which distant genomic elements contact each other (Fig. 2). Passive diffusion models are based on the assumption that the mobility of loci provides opportunities for random collisions that are then converted into productive interactions; whether they are productive is dependent on the affinity and specificity of bound protein complexes. Although these models are appealing, it seems that active processes are required, as well, to directly guide loci toward each other. For instance, enhancers have been proposed to actively track along chromatin fibers until a receptive promoter is encountered. Recently, it was found that loci can follow rapid and directed trajectories through the nucleus in an actin-dependent fashion (*19*, *20*). The roles of nuclear actin and myosin have been contentious, but these exciting recent results strongly suggest that they play critical roles in facilitating long-range interactions by transporting loci toward each other or to specific subnuclear neighborhoods, such as transcription factories, which are enriched in RNA polymerase (Fig. 1C).

Genomes also contain regulatory elements that modulate interactions between other loci. So-called insulators prevent an enhancer from activating a promoter but only when it is positioned in between them. How insulators work is not known in detail, but they too engage in long-range interactions with other elements (*21*), which suggests that they generate looped chromosome structures that somehow facilitate the formation of appropriate assemblies of enhancers and target genes.

Future Perspective

At present, significant effort is aimed at comprehensive mapping of chromosomal interactions. Several adaptations of 3C have been developed that allow large-scale detection of genomic interactions by using microarrays or by direct sequencing using any of the newly developed high-throughput sequencing technologies (*22*). The 4C method (3C-on chip, or circular 3C) allows identification of regions throughout the genome that are physically close to a single locus of interest (*23*, *24*). The 5C method (3C carbon copy) is not anchored on a single locus and is used for mapping dense interaction networks throughout large chromosomal regions of interest (*25*). These approaches will yield new insights into the spatial organization of genomes but are descriptive in nature. Additional approaches will be essential to unravel the mechanisms by which chromosomal associations affect genome regulation. These approaches include time-resolved imaging of chromosomal loci, molecular and genetic manipulation of the mechanisms that control the subnuclear localization and movement of loci, as well as biochemical studies to characterize the complexes that mediate chromosomal associations. Combined, these various approaches promise to provide exciting new insights into the three-dimensional aspects of gene regulation.

References and Notes

1. D. A. Dekker, V. van Heyningen, *Am. J. Hum. Genet.* **76**, 8 (2005).
2. A. G. West, P. Fraser, *Hum. Mol. Genet.* **14**, R101 (2005).
3. J. Dekker, K. Rippe, M. Dekker, N. Kleckner, *Science* **295**, 1306 (2002).
4. B. Tolhuis, R. J. Palstra, E. Splinter, F. Grosveld, W. de Laat, *Mol. Cell* **10**, 1453 (2002).
5. D. Vernimmen, M. De Gobbi, J. A. Sloane-Stanley, W. G. Wood, D. R. Higgs, *EMBO J.* **26**, 2041 (2007).
6. G. L. Zhou *et al.*, *Mol. Cell. Biol.* **26**, 5096 (2006).
7. C. G. Spilianakis, R. A. Flavell, *Nat. Immunol.* **5**, 1017 (2004).
8. J. A. Gross *et al.*, *Mol. Cell. Biol.* **26**, 7056 (2006).
9. H. Jing *et al.*, *Mol. Cell* **29**, 232 (2008).
10. S. Lomvardas *et al.*, *Cell* **126**, 403 (2006).
11. C. G. Spilianakis, M. D. Lalioti, T. Town, G. R. Lee, R. A. Flavell, *Nature* **435**, 637 (2005).
12. W. Xu, C. L. Tsai, J. T. Lee, *Science* **311**, 1149 (2006).
13. C. P. Bacher *et al.*, *Nat. Cell Biol.* **8**, 293 (2006).
14. S. Kurukuti *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **103**, 10684 (2006).
15. J. Dekker, *Nat. Methods* **3**, 17 (2006).
16. M. Simonis, J. Kooren, W. de Laat, *Nat. Methods* **4**, 895 (2007).
17. S. H. Fuss, M. Omura, P. Mombaerts, *Cell* **130**, 373 (2007).
18. J. D. Engel, K. Tanimoto, *Cell* **100**, 499 (2000).
19. C. H. Chuang *et al.*, *Curr. Biol.* **16**, 825 (2006).
20. M. Dundr *et al.*, *J. Cell Biol.* **179**, 1095 (2007).
21. J. A. Wallace, G. Felsenfeld, *Curr. Opin. Genet. Dev.* **17**, 400 (2007).
22. B. Wold, R. M. Meyer, *Nat. Methods* **5**, 19 (2008).
23. M. Simonis *et al.*, *Nat. Genet.* **38**, 1348 (2006).
24. Z. Zhao *et al.*, *Nat. Genet.* **38**, 1341 (2006).
25. J. Dostie *et al.*, *Genome Res.* **16**, 1299 (2006).
26. J.D. is supported by grants from NIH (HG003143), the Keck Foundation, and the Cystic Fibrosis Foundation. M. Walhout is acknowledged for suggestions for this article.

10.1126/science.1152850

PERSPECTIVE

Complex Riboswitches

Ronald R. Breaker

Using simple biochemical tricks, metabolite-binding riboswitches take on gene control functions that have long been thought to be the work of protein factors. Although modern riboswitches might be the last holdouts of primitive genetic elements, some are capable of sensory and regulatory feats that are competitive with their protein counterparts.

Riboswitches are found in mRNAs, where they bind small molecules and control gene expression (*1*). Most carry a single binding site, or aptamer, that recognizes a target ligand. The aptamer usually is located near an expression platform, a domain that alters its structure to control genes in response to metabolite binding. Riboswitches could be the most ancient of all genetic switches (*2*–*4*), but their simple construction—using only the four standard nucleotides—suggests that they might be limited in their ability to bind diverse metabolites and perform complex gene control functions. However, new insights are beginning to reveal how RNAs with potentially primitive origins and simple chemical composition can form sophisticated gene control systems.

Precision Chemical Sensing

Riboswitches must selectively recognize their targets at concentrations typically found in cells, even among a sea of competing metabolites. Discrimination against even closely related compounds is essential to avoid metabolic misregulation. Proof that RNA can achieve these goals came initially from synthetic aptamers, which have been created to bind many different ligands with high affinity and specificity (*5*). Some of these engineered RNAs function in cells as designer gene control elements (*6*), but these tend to be smaller, weaker-binding, and not as selective as natural aptamers (*4*). These differences are not surprising, given that synthetic aptamers are made in the laboratory in a few weeks, whereas the characteristics of some riboswitch aptamers have been sharpened by billions of years of evolution.

Atomic-resolution structures of riboswitch aptamers show that they make numerous hydrogen bonds with their ligands, and they form contacts that stabilize neighboring RNA interactions to further increase affinity. Some aptamers form pockets that entirely engulf the ligand, and in these instances an induced-fit mechanism of binding must occur. Also, segments of these pockets may conform precisely in shape to parts of their ligands, which permits van der Waals interactions to contribute binding energy.

The potential liability of the negative-charge character of aptamers is converted into a powerful asset by riboswitches that bind thiamine pyrophosphate (TPP) and glucosamine-6-phosphate (GlcN6P). Riboswitch classes that bind TPP (*7*, *8*) and GlcN6P (*9*, *10*) recruit positively charged magnesium ions to bridge the phosphates of the aptamer and ligand (*11*). SAM-I riboswitches bind *S*-adenosylmethionine (SAM) and discriminate against its metabolic byproduct *S*-adenosylhomocysteine by a factor of 100 (*12*). These compounds differ only by a single methyl group and a single positive charge, and it is the positive charge that gives the riboswitch its critical hold on SAM. SAM-I riboswitches use carbonyl groups (instead of phosphates) to bind this positive charge. Moreover, they exploit a bent conformation of SAM to ensure the presence of a second cationic amine group (Fig. 1).

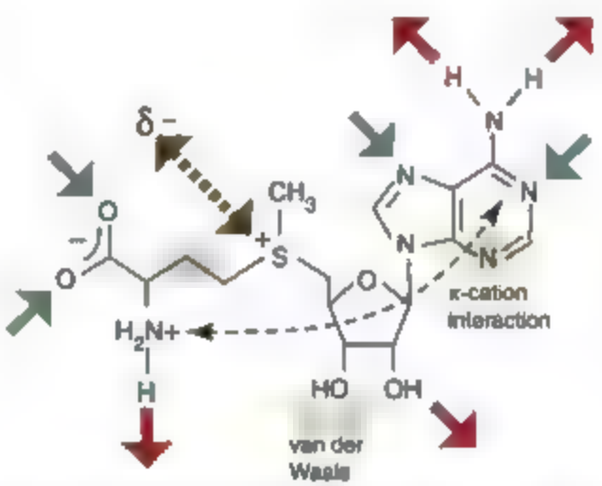


Fig. 1. Contacts made by SAM-I aptamers to selectively bind SAM (*31*). Red and green arrows designate hydrogen bond donors and acceptors, respectively. Shaded portions of SAM are not directly bound by the RNA but are required for binding, likely because their modification disrupts a π-cation interaction that stabilizes ligand conformation.

Some organisms use entirely different RNA folds to recognize SAM (*12*, *13*), and these RNAs bind SAM in distinct ways (*14*). The finding of multiple riboswitch classes for the same ligand demonstrates that RNA is a versatile medium for forming metabolite-sensing gene control elements. Thus, using large aptamers and the full range of molecular recognition options available to RNA should allow riboswitches to detect a great variety of metabolites, sometimes with more than one aptamer architecture for a given ligand.

Shape-Shifting RNAs

Riboswitches can harness the energy of ligand binding to stabilize alternatively folded forms of mRNAs. Most riboswitches have evolved to function in feedback loops, where they sense specific metabolites when they are plentiful, and shut off expression of genes for enzymes that would otherwise wastefully synthesize more (*1*). In Gram-positive bacteria, expression platforms are usually based on the mutually exclusive formation of hairpins that either permit transcription (antiterminator stems) or cause transcription termination (terminator stems) (*12*). An "OFF" switch works by binding the metabolite, preventing the antiterminator stem from forming, and permitting terminator formation. The expression platforms in Gram-negative bacteria frequently control access to ribosome binding sites (*12*), which are sometimes located within the aptamer. The latter cases are also OFF switches, because nucleotides docked in the ligand-bound aptamer cannot simultaneously play their usual role in initiation of protein synthesis. Ligand-dependent activation by riboswitches is achieved by reversing the genetic logic of the mechanisms described above.

Riboswitches that control gene expression by sensing GlcN6P do not change conformation but function as suicide RNAs (*9*, *10*). *glmS* ribozymes catalyze RNA cleavage only when bound to GlcN6P, but this special reactivity is puzzling given that all other riboswitches only modulate folding of their expression platforms. GlcN6P is positioned at the ribozyme cleavage site to serve as a cofactor, and the amine of this compound has been implicated in proton-transfer catalysis (*10*). In contrast, compounds such as glucose-6-phosphate (which lacks the amine) cannot activate ribozyme cleavage. This more sophisticated riboswitch twice discriminates against close analogs of its target, once by binding affinity differences and again by requiring the ligand to catalyze RNA cleavage.

Although riboswitch aptamers for TPP are very similar among bacteria, fungi, and plants,

Department of Molecular, Cellular and Developmental Biology, Department of Molecular Biophysics and Biochemistry, and Howard Hughes Medical Institute, Yale University, New Haven, CT 06520, USA. E-mail: ronald.breaker@yale.edu

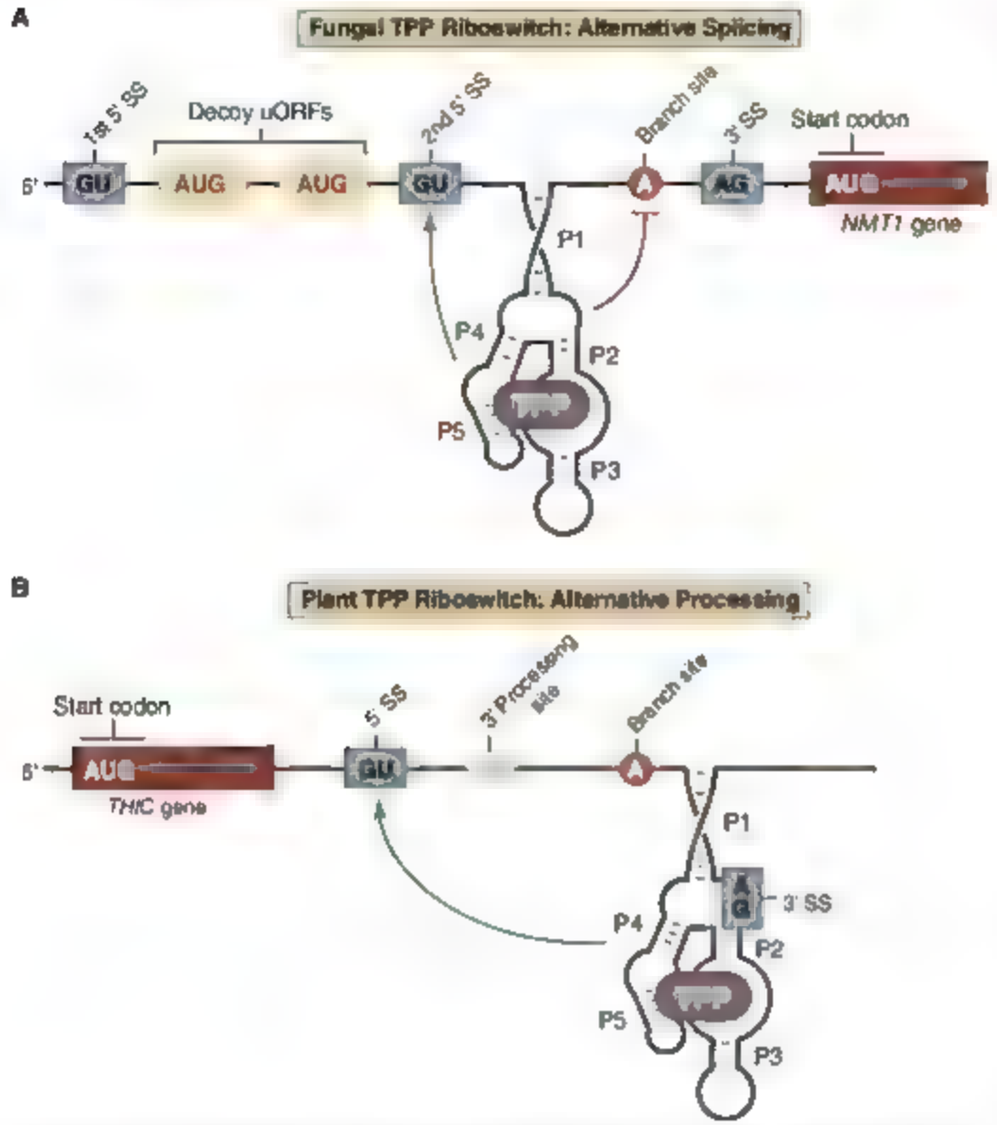


Fig. 2. Eukaryotic riboswitches that control splicing. (**A**) Riboswitch from the *NMT1* gene from the fungus *N. crassa* (*17*). (**B**) Riboswitch from the *THIC* gene from the plant *Arabidopsis thaliana*. Green and red lines identify TPP-induced folding changes that respectively activate or repress splicing elements. SS designates a splice site.

they reside in different locations in mRNAs (*15*) and they use very different gene control mechanisms (*15*–*20*). In fungi, ligand binding by some TPP riboswitches regulates access to precursor mRNA splice sites by controlling the formation of base-paired stems (Fig. 2A) (*17*). Two TPP riboswitches in the fungus *Neurospora crassa* control splicing within the 5′ untranslated regions (5′ UTRs) of mRNAs, where TPP-modulated alternative splicing dictates whether short upstream open reading frames (uORFs) are retained. These uORFs act as decoys for ribosomes and repress expression of the main ORF. TPP riboswitches in multicellular plants control splicing in the 3′ UTRs of mRNAs, which couples metabolite binding to mRNA processing and stability (*18*, *19*). This modularity of aptamers allows them to be integrated with diverse gene control mechanisms at different sites within mRNAs of organisms from all three domains of life (*15*).

Stacking Riboswitches

Any gene controlled only by a riboswitch carrying a single aptamer and expression platform requires a factor of 81 change in ligand concentration to progress from 10% to 90% gene expression (*21*). Cells using this standard configuration therefore must tolerate wide swings in the concentrations of some metabolites. However, riboswitches use at least two strategies to shrink this dynamic range. First, two or more riboswitches can be placed in series that sense the same ligand. If these tandem riboswitches are triggered with near-identical ligand concentrations, and if triggering either one independently represses gene expression, then the system requires only a factor of 40 change in ligand concentration to elicit the same change in gene expression. Adding a third riboswitch improves the responsiveness only modestly, and therefore the most common tandem arrangement observed consists of two successive riboswitches (*22*). Tandem arrangements of tRNA-sensing gene control elements called T-box RNAs are also common (*23*), most likely to increase the responsiveness of gene control to smaller changes in tRNA concentration.

A more elegant strategy to reduce dynamic range integrates multiple aptamers within a single riboswitch. This arrangement is found with many glycine riboswitches that carry two similar glycine aptamers and one expression platform (*12*, *24*). If perfectly cooperative, where ligand binding at either site improves affinity at the other, then the riboswitch could traverse the 10% to 90% gene expression range with only a factor of 9 change in glycine concentration. These tandem architectures allow RNA to compete with protein factors to create more digital genetic switches that respond to smaller changes in metabolite concentration.

Stacking riboswitches with different ligand specificities allows cells to control the same gene with multiple chemical inputs using only RNA (*22*, *25*). In *Bacillus clausii*, the methionine biosynthesis gene *metE* is controlled by tandem riboswitches for SAM and coenzyme B_{12} (*22*) that together operate as a two-input logic gate. Because SAM is made from methionine, the *metE* mRNA carries a SAM riboswitch to deactivate expression when the coenzyme is plentiful. When ample concentrations of coenzyme B_{12} are present, the B_{12} riboswitch represses *metE* because the bacterium codes for a more efficient methionine biosynthesis protein that uses a derivative of coenzyme B_{12}. Although tandem riboswitches appear infrequently in bacteria, unique riboswitch architectures can be pressed into service when cells have more challenging gene control tasks.

Kinetics Versus Thermodynamics

Riboswitches use three-dimensional folding to form a diverse array of aptamers and expression platforms, but it is the fourth dimension, or time, that adds additional and useful complexity. Ligand binding by aptamers is usually evaluated by establishing their dissociation constant (K_D) values. K_D values measured for some riboswitch aptamers are in the mid-picomolar range, and therefore riboswitches can be used to detect metabolites that are exceedingly low in concentration. Indeed, some riboswitches appear to have sufficient time to equilibrate with their surroundings and respond to concentrations of ligand that match the K_D for their aptamer (*26*). However, some riboswitches that control transcription termination operate under extreme time constraints that are too brief to allow aptamers to fully fold and equilibrate (*27*, *28*) as they are synthesized.

In these cases, it is the speed of ligand binding and the speed of action of RNA polymerase that determine what metabolite concentration will regulate gene expression. This can be an advantage for riboswitches in their evolutionary competition with protein factors. Cells can change the concentration set point for a metabolite by mutating the aptamer or protein receptor to strengthen or weaken binding. With kinetically driven riboswitches, cells can tune the set point by orders of magnitude just by accruing mutations that alter the speed of riboswitch folding, or that alter the speed with which RNA polymerase completes riboswitch synthesis.

Riboswitch Discovery

Growing DNA sequence databases and improving bioinformatics tools are rapidly expanding the number of discovered riboswitch classes in bacteria (*12*, *29*, *30*), and similar efforts in eukaryotes might also begin to uncover more riboswitches in these organisms. Currently, few eukaryotic riboswitches have been discovered, and in the evolutionary race for dominance, perhaps proteins have put more distance between themselves and their ancient RNA counterparts in this branch of the evolutionary tree. Given that eukaryotes appear to express many noncoding RNAs, researchers will need to keep open the possibility that riboswitches might also control the production and actions of these RNAs as well. As is the case with bacteria, if a protein factor cannot be found for a signaling compound, perhaps a complex sensor made of RNA is in action.

References and Notes

1. R. L. Coppins, K. B. Hall, E. A. Groisman, *Curr. Opin. Microbiol.* **10**, 176 (2007)
2. A. Nahvi *et al.*, *Chem. Biol.* **9**, 1043 (2002)
3. A. G. Vitreschak, D. A. Rodionov, A. A. Mironov, M. S. Gelfand, *Trends Genet.* **20**, 44 (2004)
4. R. R. Breaker, in *The RNA World*, R. F. Gesteland, T. R. Cech, J. F. Atkins, Eds. (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, ed. 3, 2006), pp. 89–107
5. S. E. Osborne, A. D. Ellington, *Chem. Rev.* **97**, 349 (1997)
6. J. P. Gallivan, *Curr. Opin. Chem. Biol.* **11**, 612 (2007)
7. S. Thore, M. Leibundgut, N. Ban, *Science* **312**, 1208 (2006); published online 3 May 2006 (10.1126/science.1124451)
8. A. Serganov, A. Polonskaia, A. T. Phan, R. R. Breaker, D. J. Patel, *Nature* **441**, 1167 (2006)
9. D. J. Klein, A. R. Ferré-D'Amaré, *Science* **313**, 1752 (2006)
10. J. C. Cochrane, S. V. Lipchock, S. A. Strobel, *Chem. Biol.* **14**, 97 (2007)
11. T. E. Edwards, D. J. Klein, A. R. Ferré-D'Amaré, *Curr. Opin. Struct. Biol.* **17**, 273 (2007)
12. J. E. Barrick, R. R. Breaker, *Genome Biol.* **8**, R239 (2007)
13. R. T. Fuchs, F. J. Grundy, T. M. Henkin, *Nat. Struct. Mol. Biol.* **13**, 226 (2006)
14. S. D. Gilbert, R. P. Rambo, D. Van Tyne, R. T. Batey, *Nat. Struct. Mol. Biol.* **15**, 177 (2008)
15. N. Sudarsan, J. E. Barrick, R. R. Breaker, *RNA* **9**, 949 (2003)
16. F. Kubodera *et al.*, *FEBS Lett.* **555**, 516 (2003)
17. M. T. Cheah, A. Wachter, N. Sudarsan, R. R. Breaker, *Nature* **447**, 497 (2007)
18. S. Bocobza *et al.*, *Genes Dev.* **21**, 2874 (2007)
19. A. Wachter *et al.*, *Plant Cell* **19**, 3437 (2007)
20. M. T. Croft, M. Moulin, M. E. Webb, A. G. Smith, *Proc. Natl. Acad. Sci. U.S.A.* **104**, 20770 (2007)
21. R. Welz, R. R. Breaker, *RNA* **13**, 573 (2007)
22. N. Sudarsan *et al.*, *Science* **314**, 300 (2006)
23. A. Gutiérrez-Preciado, R. A. Jensen, C. Yanofsky, E. Merino, *Trends Genet.* **21**, 492 (2005)
24. M. Mandal *et al.*, *Science* **306**, 275 (2004)
25. C. D. Stoddard, R. T. Batey, *ACS Chem. Biol.* **1**, 751 (2006)
26. R. Rieder, K. Lang, D. Graber, R. Micura, *ChemBioChem* **8**, 896 (2007)
27. J. K. Wickiser, W. C. Winkler, R. R. Breaker, D. M. Crothers, *Mol. Cell* **18**, 49 (2005)
28. W. J. Greenleaf, K. L. Frieda, D. A. N. Foster, M. T. Woodside, S. M. Block, *Science* **319**, 630 (2008); published online 2 January 2008 (10.1126/science.1151298)
29. Z. Weinberg *et al.*, *Nucleic Acids Res.* **35**, 4809 (2007)
30. M. D. Kazanov, A. G. Vitreschak, M. S. Gelfand, *BMC Genomics* **8**, 347 (2007)
31. R. K. Montange, R. T. Batey, *Nature* **441**, 1172 (2006)

32. Riboswitch research in the Breaker laboratory is supported by grants from NIH and the Howard Hughes Medical Institute.

10.1126/science.1152621

PERSPECTIVE

Evolution of Eukaryotic Transcription Circuits

Brian B. Tuch,[1,2] Hao Li,[1] Alexander D. Johnson[1,2]*

The gradual modification of transcription circuits over evolutionary time scales is an important source of the diversity of life. Over the past decade, studies in animals have shown how seemingly small molecular changes in gene regulation can have large effects on morphology and physiology and how selective pressures can act on these changes. More recently, genome-wide studies, particularly those in single-cell yeasts, have uncovered evidence of extensive transcriptional rewiring, indicating that even closely related organisms regulate their genes using markedly different circuitries.

Transcription of each gene in a eukaryotic organism is controlled by a collection of cis-regulatory sequences that are typically positioned in proximity to the coding sequence. The collection of cis-regulatory sequences associated with each gene specifies the time and place in the organism that the gene is to be transcribed. This information is read by sequence-specific DNA binding proteins [herein called transcription regulators (TRs)], which recognize these sequences and which themselves are typically expressed or active only at particular times and places in the life of the organism. It is the combination of active TRs present at a particular location and time that selects, through interactions with cis-regulatory sequences, those genes to be transcribed. Of course, there are many additional steps in transcription and in gene regulation; nevertheless, the cis-regulatory sequences and the TRs that recognize them form a critical layer of gene regulation.

Several properties of transcription regulation are especially important for considering its role in evolution (*1*–*4*). cis-regulatory sequences are short (generally 5 to 10 nucleotides), degenerate (similar sequences confer equivalent TR binding), and their positions, relative to the gene whose transcription they control, can be variable. Different cis-regulatory sequences are often found close to one another, and TRs often bind cooperatively to these adjacent sites. This cooperative binding is a form of combinatorial control: the use of multiple, rather than single, TRs to control expression of a gene. cis-regulatory sequences often cluster into modules, each module acting independently to direct expression of the gene to a particular part of the organism at a specified time.

TRs are also modular and, in the laboratory, bits and pieces from different TRs can be recombined to produce novel types of regulation. Mutations can alter their DNA-binding specificity, their partner proteins, and their influence (activating or repressing) on transcription. Because many of the crucial protein-protein interactions made by TRs are relatively weak and nonspecific, even small changes to them can have large effects on gene regulation.

These properties make it easy to understand how new patterns of gene expression could arise through simple mutations. During the past decade, studies of single genes in animals have revealed many striking cases in which changes in cis-regulatory sequences likely underlie new morphological or physiological features (*5*, *6*). These include the evolution of lactase persistence in humans (*7*), bone structures in fish (*8*), and trichomes (*9*) and pigmentation (*10*) in flies.

[1]Department of Biochemistry and Biophysics, University of California, San Francisco, CA, USA. [2]Department of Microbiology and Immunology, University of California, San Francisco, CA, USA.

*To whom correspondence should be addressed. E-mail: ajohnson@cgl.ucsf.edu

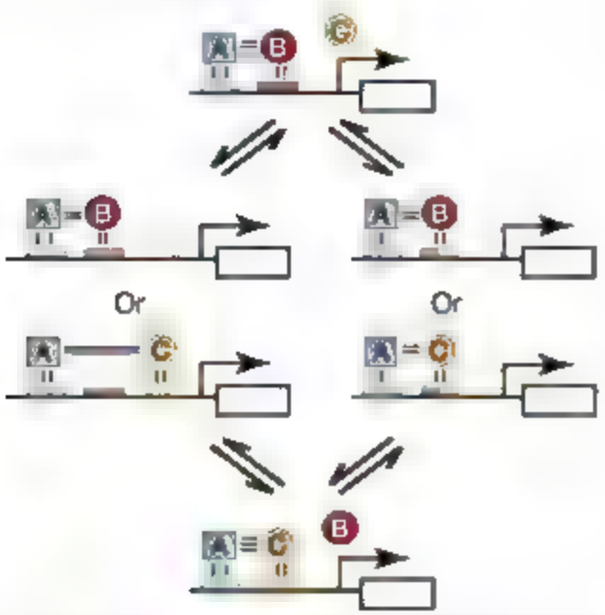


Fig. 1. Pathways to the rewiring of combinatorial circuitry. These two schemes can account for a handoff in the control of a gene (or a set of genes) from one TR to another. In both pathways an intermediate stage exists in which regulators B and C may act redundantly. Small black lines represent protein-protein and protein-DNA interactions, the number of these indicating the strength of the interaction. At any given time, each gene within a co-expressed set may have different control states (B only, C only, or B and C). The left pathway may be the route by which ribosomal genes and galactose-metabolizing genes were rewired in fungi (*14*, *19*). The right pathway is the likely route by which a-specific genes were rewired (*20*).

Although the emphasis is often placed on cis-regulatory sequences, changes to TRs can also underlie phenotypic change (*11*, *12*).

The animal studies outlined above typically start with an observable intra- or interspecies difference and trace its origins to changes in gene regulation at an individual locus. A complementary approach to studying evolutionary changes in transcriptional regulation (rewiring) begins with a molecular description of a transcription circuit, typically a large one consisting of several TRs and many target genes (i.e., genes they bind to and regulate). The circuit is then compared among two or more species. An advantage of this strategy is that it allows the entire landscape of circuit rewiring to be surveyed, without any bias as to the consequences of rewiring events to the organism. Of course, this is also its principal limitation; it is often difficult to discern whether the changes observed provide (or provided in the past) any benefit to the organism.

This genomic approach has been used to compare circuitry in closely related yeast, fly, and mammal species. Typically, bioinformatics, transcriptional profiling, and full-genome chromatin immunoprecipitation are used, often in combination. This approach has provided support for previous ideas and has also produced new insights.

First, high levels of transcriptional rewiring can occur over relatively short evolutionary time scales (*13*, *14*). Although the DNA-binding specificities of orthologous TRs (that is, TRs related by direct descent from the last common ancestor of a group of species) rarely differ substantially across species, the genes they directly regulate can differ considerably. For example, comparisons of the binding profiles of four liver-specific TRs across 4000 genes in mouse and human hepatocytes found that less than two-thirds of genes are conserved as targets of each TR (*15*). A genome-wide study of two TRs in three closely related yeast species (20 million years divergence) estimated that only a third of the TR target gene connections seen in one species were preserved in the other two (*16*). Although some of these differences could be attributed to loss and gain of cis-regulatory sequences, others could not, and it remains to be seen what other types of molecular changes (e.g., changes in the activity level of TRs) contribute to this divergence.

A study that examined combinatorial circuitry involving the TR Mcm1 and its cofactors across three highly divergent yeasts (~300 million years divergence) also found evidence of massive rewiring (*17*). Only about 15% of the direct Mcm1 target gene interactions of *Saccharomyces cerevisiae* were preserved in two other yeast species. Mcm1 binds cooperatively to DNA with a set of cofactors to regulate many genes in each species, and the extensive rewiring observed was traced to high rates of gain and loss of cis-regulatory sequences as well as to the formation of new Mcm1-cofactor combinations and the breaking of old ones.

Second, genomic approaches have shown that the same set of coexpressed genes can be regulated by different mechanisms in different species. Earlier studies in flies showed that stabilizing selection can maintain the expression pattern of a single gene, while still allowing for considerable drift in the underlying regulatory mechanism [e.g., (*18*)]. Genome-wide studies in yeast have extended this idea, uncovering examples in which an entire group of genes remain coexpressed (that is, the genes respond as a group to changes in the environment or other perturbations) in different species, but the TR responsible for the regulation seems to have been swapped in one species relative to another. For example, in *S. cerevisiae* the presence of galactose induces transcription of genes that produce galactose-metabolizing enzymes via the TR Gal4. In another yeast, *Candida albicans*, the same enzymes are induced by galactose, but the Gal4 ortholog seems to have no role in this regulation, instead, these genes appear to be controlled by cis-regulatory sequences recognized by a different TR, and the Gal4 ortholog regulates glycolytic enzymes (*19*).

Another example occurs in mating-type regulation in fungi. In the lineage leading to *S. cerevisiae*, regulation of the coexpressed a-specific genes (transcribed in a cells and not in α cells) was "handed off" from a transcriptional activator to a transcriptional repressor (*20*). Because the activator and repressor are expressed in opposite cell types, the overall logic of the circuit is conserved. These replacements, of one TR with another, likely occurred through an intermediate state in which the target genes came under dual regulation, thus preserving co-expression throughout the transition (Fig. 1). Transition through a redundant intermediate has also been suggested for changes in the regulation of ribosomal genes in fungi (*14*).

It is not yet clear whether the rewiring of these coexpressed gene sets provides any advantage to the organism, as the overall expression pattern of the target genes seems, at least superficially, to have remained constant. It is possible that many examples of transcriptional rewiring are not adaptive at all but may simply reflect neutral evolution between alternative regulatory schemes (*21*).

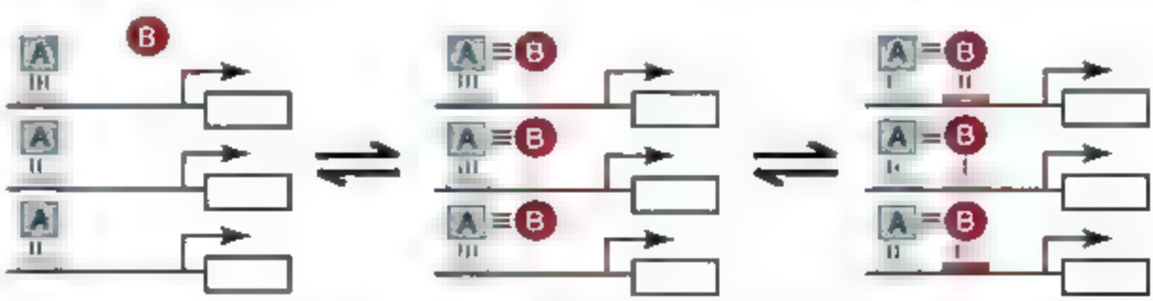


Fig. 2. A plausible pathway to the concurrent rewiring of a large set of genes. In this scenario an interaction is acquired between TRs A and B, after which interactions between B and DNA are optimized gene-by-gene. Rewiring in this manner could avoid fitness barriers imposed by initially changing regulation one gene at a time.

Finally, genomic approaches provide evidence consistent with the idea that cooperative binding of TRs facilitates circuit changes. In its simplest form, the occupancy of two cooperatively binding TRs, A and B, on DNA depends on the concentration of each protein, the strength of each protein-DNA interaction, and the net favorable interaction between the two proteins. A decrease in any one of these parameters can be compensated by a gain in any other. This allows substantial shifts in the relative contribution of each component to the overall energetics without destroying the regulation; this flexibility, in turn, could catalyze regulatory change. For example, as shown in Fig. 1 (right path), the cis-regulatory sequence of B could drift away from consensus if the A-B interaction were sufficiently favorable. This drift could produce a weak cis-regulatory sequence for a third TR, C, whose expression might overlap that of B. If the A-C interaction were then strengthened by point mutation, the regulation of the gene would have changed from A-B to A-C through a series of small steps, none of which would destroy regulation of the gene. This scenario is but one of many that is made possible by cooperative binding. If the number of cooperative components is increased, the possibilities for "movement" in the system are multiplied.

A few studies provide direct support for these ideas. For example, the fungal mating circuit change already described roughly follows the scenario presented above (*20*). Further evidence comes from a whole-network analysis of the transcriptional circuitry of *S. cerevisiae* (*22*). Here, a strong correlation was observed between the number of TRs that regulate a gene and the fuzziness (departure from consensus) of the cis-regulatory sequences present at that gene. This fuzziness may indicate that the cooperative binding of multiple TRs to DNA relaxes the importance of any one TR-DNA interaction. It has also been shown, through the simulated evolution of systems of interacting components, that the existence of redundant intermediate states, such as those described above, greatly catalyzes change within these systems (*23*). Finally, Zuckerkandl has argued that the type of neutral changes permitted by cooperative assembly of TRs on DNA may have facilitated the formation of complex regulatory circuits (*24*). Although we have emphasized cooperative binding of TRs to DNA, other forms of combinational control (e.g., when two TRs bind DNA independently to control a target gene) could also facilitate circuit rewiring.

We propose that cooperativity may be especially important for coordinating changes in the regulation of entire sets of coexpressed genes. For example, the gain of a protein-protein interaction between two TRs could "jump-start" the rewiring of a set of genes at which one TR is already present (Fig. 2). Afterward, the new circuit could be improved, target gene by target gene, through the gradual formation of optimal cis-regulatory sequences. This idea may help to explain how regulatory changes could sweep through a complete set of coexpressed genes.

References and Notes

1. S. B. Carroll, J. K. Grenier, S. D. Weatherbee, *From DNA to Diversity: Molecular Genetics and the Evolution of Animal Design* (Blackwell, Malden, MA, ed. 2, 2005).
2. E. H. Davidson, *Genomic Regulatory Systems: Development and Evolution* (Academic Press, San Diego, 2001).
3. G. A. Wray *et al.*, *Mol. Biol. Evol.* **20**, 1377 (2003).
4. M. Kirschner, J. Gerhart, *Proc. Natl. Acad. Sci. U.S.A.* **95**, 8420 (1998).
5. G. A. Wray, *Nat. Rev. Genet.* **8**, 206 (2007).
6. B. Prud'homme, N. Gompel, S. B. Carroll, *Proc. Natl. Acad. Sci. U.S.A.* **104** (suppl. 1), 8605 (2007).
7. S. A. Tishkoff *et al.*, *Nat. Genet.* **39**, 31 (2007).
8. M. D. Shapiro *et al.*, *Nature* **428**, 717 (2004).
9. A. P. McGregor *et al.*, *Nature* **448**, 587 (2007).
10. S. Jeong *et al.*, *Cell* **132**, 783 (2008).
11. M. Ronshaugen, N. McGinnis, W. McGinnis, *Nature* **415**, 914 (2002).
12. R. Galant, S. B. Carroll, *Nature* **415**, 910 (2002).
13. A. P. Gasch *et al.*, *PLoS Biol.* **2**, e398 (2004).
14. A. Tanay, A. Regev, R. Shamir, *Proc. Natl. Acad. Sci. U.S.A.* **102**, 7203 (2005).
15. D. T. Odom *et al.*, *Nat. Genet.* **39**, 730 (2007).
16. A. R. Borneman *et al.*, *Science* **317**, 815 (2007).
17. B. B. Tuch, D. J. Galgoczy, A. D. Hernday, H. Li, A. D. Johnson, *PLoS Biol.* **6**, e38 (2008).
18. M. Z. Ludwig, C. Bergman, N. H. Patel, M. Kreitman, *Nature* **403**, 564 (2000).
19. M. Martchenko, A. Levitin, H. Hogues, A. Nantel, M. Whiteway, *Curr. Biol.* **17**, 1007 (2007).
20. A. E. Tsong, B. B. Tuch, H. Li, A. D. Johnson, *Nature* **443**, 415 (2006).
21. M. Lynch, *Nat. Rev. Genet.* **8**, 803 (2007).
22. Y. Bilu, N. Barkai, *Genome Biol.* **6**, R103 (2005).
23. E. S. Haag, M. N. Molla, *Evol. Int. J. Org. Evol.* **59**, 1620 (2005).
24. E. Zuckerkandl, *J. Mol. Evol.* **44** (suppl. 1), S2 (1997).

25. We thank B. Alberts, S. Carroll, A. Gann, S. Noble, and M. Plasterer for comments on the manuscript. We regret that due to space limitations we could not include original citations to much of the work in this field. This work was supported by grants from the NIH (to A.D.J. and H.L.) and the Packard Fellowship in Science and Engineering (to H.L.). B.B.T. was supported by an NSF Predoctoral Fellowship.

10.1126/science.1152398

Dynamics of Saturn's South Polar Vortex

Ulyana A. Dyudina,[1]* Andrew P. Ingersoll,[1] Shawn P. Ewald,[2] Ashwin R. Vasavada,[2] Robert A. West,[2] Anthony D. Del Genio,[3] John M. Barbara,[3] Carolyn C. Porco,[4] Richard K. Achterberg,[5] F. Michael Flasar,[5] Amy A. Simon-Miller,[5] Leigh N. Fletcher[2,6]

Most planets with an atmosphere have large vortices. Here, we present observations of Saturn's south polar vortex (SPV) showing that it has a unique combination of properties, resembling some vortices in some respects but not any other vortex in all respects.

Our data are from observations over 3 hours by Cassini on 11 October 2006. A false-color image of cloud heights (Fig. 1A) shows a dark, red, central eye, indicating a nearly cloud-free upper atmosphere above lower, tropospheric clouds. The blue-green ring outside the eye indicates high clouds and haze, which is consistent with uplifted air. The eye has two concentric boundaries.

The eye wall clouds cast shadows that followed the sun in a counterclockwise direction as the planet turned (Fig. 1, B to D). From the shadow lengths, we estimate that the outer wall is 40 ± 20 km high and that the inner wall is 70 ± 30 km high, about twice the pressure scale height of Saturn's atmosphere. The eye wall clouds seem to extend up to the tropopause, which is at the ~100 mbar level (*1*).

We tracked the motion of individual cloud features (*2*). The peak zonal velocity, $\bar{u}$, was 150 ± 20 m s^{-1} near the outer eye wall. Absolute vorticity consists of two parts: a part ζ due to motion relative to the planet and a part f due to the planet's rotation (*3*). Up to latitude 85°, the measured $\bar{u}$ increased slightly faster than a constant absolute vorticity profile. Poleward of 85°, $\bar{u}$ increased more slowly. Constant absolute vorticity is consistent with horizontal stirring by eddies. The angular momentum in Saturn's SPV decreased toward its center. We observed no poleward or equatorward mean motion.

The relative vorticity, ζ, estimated from the measured $\bar{u}$ is close to zero outside the outer eye wall. The puffy red clouds in Fig. 1A are anticyclones (*2*), with a vorticity of $-1 \times 10^{-4} \pm 1 \times 10^{-4}$ s^{-1}, which is −1/3 the magnitude of the planetary vorticity, f, but of opposite sign. This result is consistent with a convective origin because parcels rising from the convective interior should have $\zeta + f = 0$ when they spread out in the upper troposphere (*3*).

Fig. 1. Images of Saturn's south polar clouds taken by the Cassini imaging science subsystem (ISS). The images have been map-projected with use of polar stereographic projection. Latitudes are planetocentric. **(A)** False-color image from light at 889 nm, 727 nm, and 750 nm projected as blue, green, and red, respectively. In the original images, sunlight was attenuated by a factor of e (2.71...) at the 80-mbar and 300-mbar levels for light at 889 and 727 nm, respectively. Sunlight passes through to deeper levels for light at 750 nm. Thus clouds below 300 mbar appear red, and high thin clouds appear blue or green (*2*). The eye walls can be seen in all three color planes and thus extend above the 80-mbar surface. **(B to D)** Time sequence showing shadows (the dark crescent-shaped areas inside the walls). The first map was taken on 11 October 2006 at 19 hours 42 min. The maps are labeled by the time lapsed since the first map. The white arrow shows the direction of propagation of the incident sunlight.

Cassini Composite Infrared Spectrometer (CIRS) data show that the vortex is anomalously warm, particularly just beneath the tropopause (by 5 K) but also in the stratosphere (by 3 to 4 K) (*1*). The warm central core means that the central low pressure, and with it the cyclonic circulation, should weaken with altitude if the flow is balanced. However, a movie of images like Fig. 1A shows no weakening (*2*).

The failure of the wind to weaken means that the centrifugal force at high altitudes is not completely balanced by the inward pressure force. This unbalanced force could drive an outward flow.

The SPV is a warm-core feature with cyclonic (clockwise) relative vorticity. Like a terrestrial hurricane, it has an eye, eye wall clouds, and multiple convective clouds outside the eye. However, hurricanes exist in the tropics, are not stationary, and derive their energy from interaction with the underlying ocean. The SPV is different from Jupiter's Great Red Spot and white ovals, which are anticyclones with uniformly high clouds at their centers (*4*). Observations do not cover the poles of Jupiter well enough to detect a possible vortex there.

In some respects, the SPV resembles the polar vortices on Venus, which are cyclonic and have warm features at the poles, although the Venus features are dipole-shaped, have cold collars, and are not surrounded by convective clouds (*5*). Similar to Saturn's atmosphere, Neptune's atmosphere is warm poleward of 70° at altitudes near 100 mbar (*6*). The SPV is different from Earth's Arctic and Antarctic polar vortices, which are cold-core features that are not associated with clouds and/or convection.

References and Notes

1. L. N. Fletcher *et al.*, *Science* **319**, 79 (2008).
2. Materials and methods are available on *Science* Online.
3. J. R. Holton, *An Introduction to Dynamic Meteorology* (Elsevier, Amsterdam, ed. 4, 2004).
4. F. Bagenal, Ed., *Jupiter: The Planet, Satellites and Magnetosphere* (Oxford Univ. Press, Oxford, 2003).
5. G. Piccioni *et al.*, *Nature* **450**, 637 (2007).
6. G. S. Orton, T. Encrenaz, C. Leyrat, R. Puetter, A. J. Friedson, *Astron. Astrophys.* **473**, L5 (2007).
7. This research was supported by the NASA Cassini Project.

Supporting Online Material

www.sciencemag.org/cgi/content/full/319/5871/1801/DC1
Materials and Methods
Fig. S1
References
Movie S1

30 November 2007; accepted 25 January 2008
10.1126/science.1153633

[1]Division of Geological and Planetary Sciences, California Institute of Technology, Pasadena, CA 91125, USA. [2]Jet Propulsion Laboratory, California Institute of Technology, Pasadena, CA 91109, USA. [3]Goddard Institute for Space Studies, NASA, 2880 Broadway, New York, NY 10025, USA. [4]Space Science Institute, 4750 Walnut Street, Suite 205, Boulder, CO 80301, USA. [5]NASA Goddard Space Flight Center, Code 693, Greenbelt, MD 20771, USA. [6]Department of Physics, Clarendon Laboratory, University of Oxford, Parks Road, Oxford OX1 3PU, UK.

*To whom correspondence should be addressed. E-mail: ulyana@gps.caltech.edu

Magnetar-Like Emission from the Young Pulsar in Kes 75

F. P. Gavriil,[1,2]* M. E. Gonzalez,[3] E. V. Gotthelf,[4] V. M. Kaspi,[3] M. A. Livingstone,[3] P. M. Woods[5,6]

We report the detection of magnetar-like x-ray bursts from the young pulsar PSR J1846–0258, at the center of the supernova remnant Kes 75. This pulsar, long thought to be exclusively rotation-powered, has an inferred surface dipolar magnetic field of 4.9×10^{13} gauss, which is higher than those of the vast majority of rotation-powered pulsars, but lower than those of the approximately 12 previously identified magnetars. The bursts were accompanied by a sudden flux increase and an unprecedented change in timing behavior. These phenomena lower the magnetic and rotational thresholds associated with magnetar-like behavior and suggest that in neutron stars there exists a continuum of magnetic activity that increases with inferred magnetic field strength.

Magnetars are young isolated neutron stars with ultrahigh magnetic fields (*1*, *2*). Observational manifestations of these exotic objects include the soft gamma repeaters (SGRs) and the anomalous x-ray pulsars (AXPs). Magnetars exhibit a variety of forms of radiative variability unique to their source class; these include short (<1 s) x-ray and gamma-ray bursts, and sudden flux enhancements that decay on time scales of weeks to months, both of which are too bright to be powered by rotational energy loss (*3*). A major puzzle in neutron star physics has been what distinguishes magnetars from neutron stars that have comparably high fields yet no apparent magnetar-like emission (*4*).

The 326-ms pulsar PSR J1846–0258 is the central isolated neutron star associated with the young shell-type supernova remnant (SNR) Kes 75 [SNR G29.6+0.1, see (*5*) for details]. If standard magnetic dipole braking is assumed, this pulsar has among the largest dipolar magnetic fields of the known young rotation-powered pulsars and the sixth largest overall, $B \equiv 3.2 \times 10^{19}\,\mathrm{G}\sqrt{P\dot{P}} = 4.9 \times 10^{13}$ G, where P is the period (in seconds), and $\dot{P}$ is the period derivative. In addition, its spin-down age of $\tau \equiv P/(n-1)\dot{P} = 884$ years is the smallest of all known pulsars (*5*, *6*). The observed x-ray luminosity of PSR J1846–0258 is $L = 4.1 \times 10^{34}(d/6\ \mathrm{kpc})^2$ erg s^{-1} in the 3- to 10-keV band, for a distance of $d \sim 6$ kpc, the mean distance found from HI and ^{13}CO spectral measurements (*7*). The pulsar has all the hallmarks of being rotation-powered, with a radiative output well under its spin-down luminosity ($\dot{E} \equiv 3.9 \times 10^{36}\,\dot{P}/P^3$ erg s^{-1} 8.1×10^{36} erg s^{-1}), an otherwise unremarkable braking index ($n \sim 2.65$) (*6*), and a bright pulsar wind nebula (Fig. 1). This pulsar is one of only about three young rotation-powered pulsars from which no radio emission is detected (*8*), although this may be due to beaming.

Observations in the direction of Kes 75 obtained with the Rossi X-ray Timing Explorer (RXTE) have revealed several short bursts of cosmic origin lasting <0.1 s (Fig. 2). We discovered four bursts in a 3.4-ks observation made on 31 May 2006 and a fifth in a 3.5-ks observation made on 27 July 2006.

These data were obtained with the Proportional Counter Array (PCA) onboard RXTE, which provides approximately microsecond time resolution and 256 spectral channels over the ~2- to 60-keV bandpass and consists of five independent proportional counter units (PCUs). The bursts are plotted in Fig. 2 and their properties are listed in Table 1. We quantified the burst properties as we have for those seen in bursting AXPs [supporting online material (SOM) text and (*9*–*12*)]. All five bursts were highly significant and were recorded in all operational PCUs simultaneously. We found no additional bursts in the 21.4 Ms of available data on this field collected by RXTE over the past 7 years.

Because of the PCA's large (1° × 1°) field of view, the origin of the bursts was not immediately apparent. However, we could unambiguously identify PSR J1846–0258 as their origin because the bursts coincided with a dramatic rise in its pulsed flux, which lasted ~2 months (Fig. 2) and was remarkably similar to those observed from AXPs (*13*–*15*). The pulsed flux was extracted according to the method detailed in (*15*) and corrected for collimator response and exposure for each PCU. We modeled the recovery from the pulsed flux enhancement as an exponential decay (with a $1/e$ time constant of 55.5 ± 5.7 days) and estimated a total 2- to 60-keV energy release of $3.8 \times 10^{41}(d/6\ \mathrm{kpc})^2$ erg to $4.8 \times 10^{41}(d/6\ \mathrm{kpc})^2$ erg, assuming isotropic emission. If we assume a power-law model for the flux decay, commonly used for the magnetars, we obtain an index of −0.63 ± 0.06. However, this model is rejected with reduced chi-squared statistic χ^2_ν [51 degrees of freedom (df)] = 1.31 as compared with χ^2_ν (51 df) = 0.95 for the exponential model.

At the onset of the outburst, the timing noise of the source changed dramatically from that typical of a young rotation-powered pulsar to that typical of AXPs. PSR J1846–0258 was spinning down smoothly with a braking index of n 2.65 ± 0.01 (*6*) until phase coherence was lost on modified Julian day (MJD) 53886, the same observation in which the first four bursts were observed. This loss of phase coherence could signal a spin-up glitch as has been seen to accom-

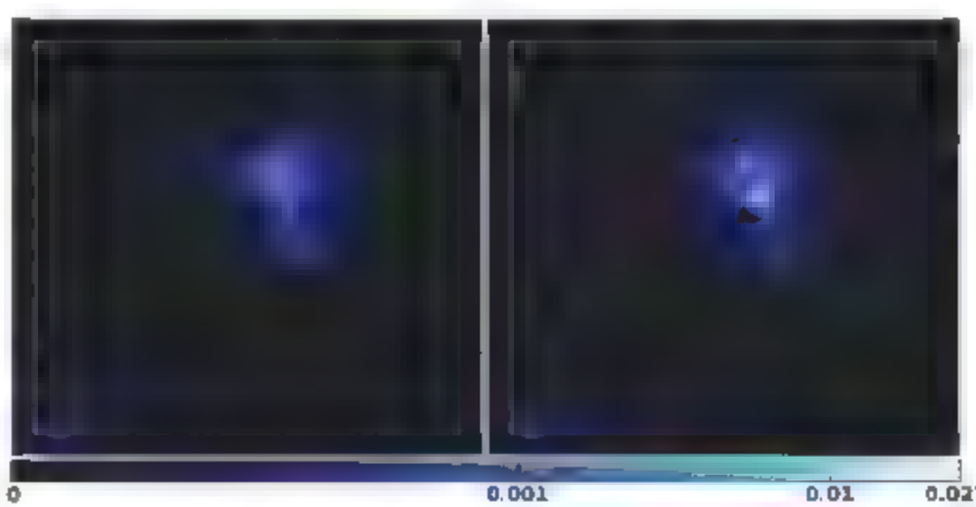


Fig. 1. High-resolution Chandra x-ray images (0.5 to 10 keV) of PSR J1846–0258 in SNR Kes 75 centered on the pulsar and its surrounding pulsar wind nebula, obtained before and during the 2006 outburst. After the bursts, the pulsar became brighter as well as softer. These images were made using archival ACIS-S3 observations obtained on 15 to 16 October 2000 (**left**) and very fortuitously on 5, 7 to 8, 9, and 12 to 13 June 2006 (**right**) and are background-subtracted, exposure-corrected, smoothed with a constant Gaussian with width $\sigma = 0.5'$, and finally displayed using the same brightness scale.

[1]NASA/Goddard Space Flight Center, Code 662, Greenbelt, MD 20771, USA. [2]Center for Research and Exploration in Space Science and Technology, University of Maryland Baltimore County, 1000 Hilltop Circle, Baltimore, MD 21250, USA. [3]Department of Physics, McGill University, 3600 University Street, Montreal, QC H3A 2T8, Canada. [4]Columbia Astrophysics Laboratory, Columbia University, 550 West 120th Street, New York, NY 10027–6001, USA. [5]Dynetics, 1000 Explorer Boulevard, Huntsville, AL 35806, USA. [6]National Space Science and Technology Center, 320 Sparkman Drive, Huntsville, AL 35805, USA.

*To whom correspondence should be addressed. E-mail: lotis.p.gavriil@nasa.gov

pany other AXP radiative events (*13*, *17*, *18*). The dramatic sudden timing noise makes the determination of accurate glitch parameters via phase-coherent timing difficult. In the most recent data, the timing noise appears to have settled somewhat, though has not relaxed to its pre-burst behavior.

We also examined archival high-resolution charge-coupled device (CCD) images of Kes 75 obtained with the Chandra X-ray Observatory (CXO) both before (October 2000) and very fortuitously during (June 2006) the event. This allowed us to identify the dramatic change in the flux of the pulsar relative to its bright, but relatively constant, pulsar wind nebula (Fig. 1 and SOM text).

The CXO-measured spectrum at the outburst epoch softened significantly relative to quiescence. A fit to a power-law model in 2006 produced a larger value for the photon index Γ with $\Gamma = 1.89^{+0.04}_{-0.06}$ and $1.17^{+0.15}_{-0.12}$ for epochs 2006 and 2000, respectively (3σ errors). The larger value of the photon index was then closer to those seen in magnetars ($\Gamma \sim 2$ to 4). Because of this softening, the 0.5- to 2-keV flux showed the largest increase, a factor of 17^{+11}_{-6} whereas the 2- to 10-keV flux increased by a factor of $5.5^{+4.5}_{-2.2}$ (3σ errors, Fig. 1). Although the 2006 spectrum is softer, the large absorption precludes the identification of any substantial thermal components. The CXO spectral analysis was non-trivial because of the brightness of the source and associated CCD pileup (see SOM text for details).

The coincidence of the bursts with the flux enhancement (Fig. 2), the distinct changes in the pulsar spectral properties (Fig. 1), and the timing anomaly and sudden change in timing noise properties all firmly establish PSR J1846–0258 as the origin of the bursts.

This is the first detection of x-ray bursts from an apparent rotation-powered pulsar. It is instructive to compare these burst properties with those of SGRs and AXPs. SGRs are characterized by their frequent hyper-Eddington ($\sim 10^{41}$ erg s^{-1}) and short ($\sim$0.1 s) repeat x-ray bursts. AXPs also emit such bursts, albeit less frequently (*9*). The bursts from PSR J1846–0258 were short (<0.1 s), showed no emission lines in their spectra, and occurred preferentially at pulse maximum (Table 1). The peak luminosities (L_p) of all bursts were greater than the Eddington luminosity (L_E) for a 1.4 solar mass neutron star, assuming isotropic emission and a distance of $d = 6$ kpc (*7*) (burst 2 had $L_p > 10\ L_E$). Considering the distribution of SGR- and AXP-burst temporal, energetic, and spectral properties (*10*, *19*), the Kes 75 bursts are indistinguishable from many of the bursts seen in AXPs and SGRs.

PSR J1846–0258's pulsed flux flare is also a magnetar hallmark. A twisted magnetosphere and associated magnetospheric currents induce enhanced surface thermal x-ray emission and resonant upscattering thereof (*20*, *21*). Flux enhancements and their subsequent decay in AXPs have been interpreted as sudden releases of energy (either above or below the crust) followed by thermal afterglow, in which case there is an abrupt rise with a gradual decay. A power-law fit was an excellent characterization of AXP 1E 2259+586's flux decay after its 2002 outburst. For PSR J1846–0258, such a model did not fit the data as well as an exponential model (Fig. 2). Spectral changes are also expected with these enhancements. The softening of the source's spectrum suggests that it underwent a transition from a purely magnetospheric-type spectrum, typical of energetic rotation-powered pulsars, to one consistent with the persistent emission from magnetars. For this reason, it is difficult to directly compare the spectral characteristics of this flux enhancement to those of other magnetars in outburst. The total 2- to 10-keV energy released during the flux enhancement [$3.3 \times 10^{41} (d/6\ \mathrm{kpc})^2$ erg to $3.8 \times 10^{41} (d/6\ \mathrm{kpc})^2$ erg, assuming isotropic emission] is comparable to

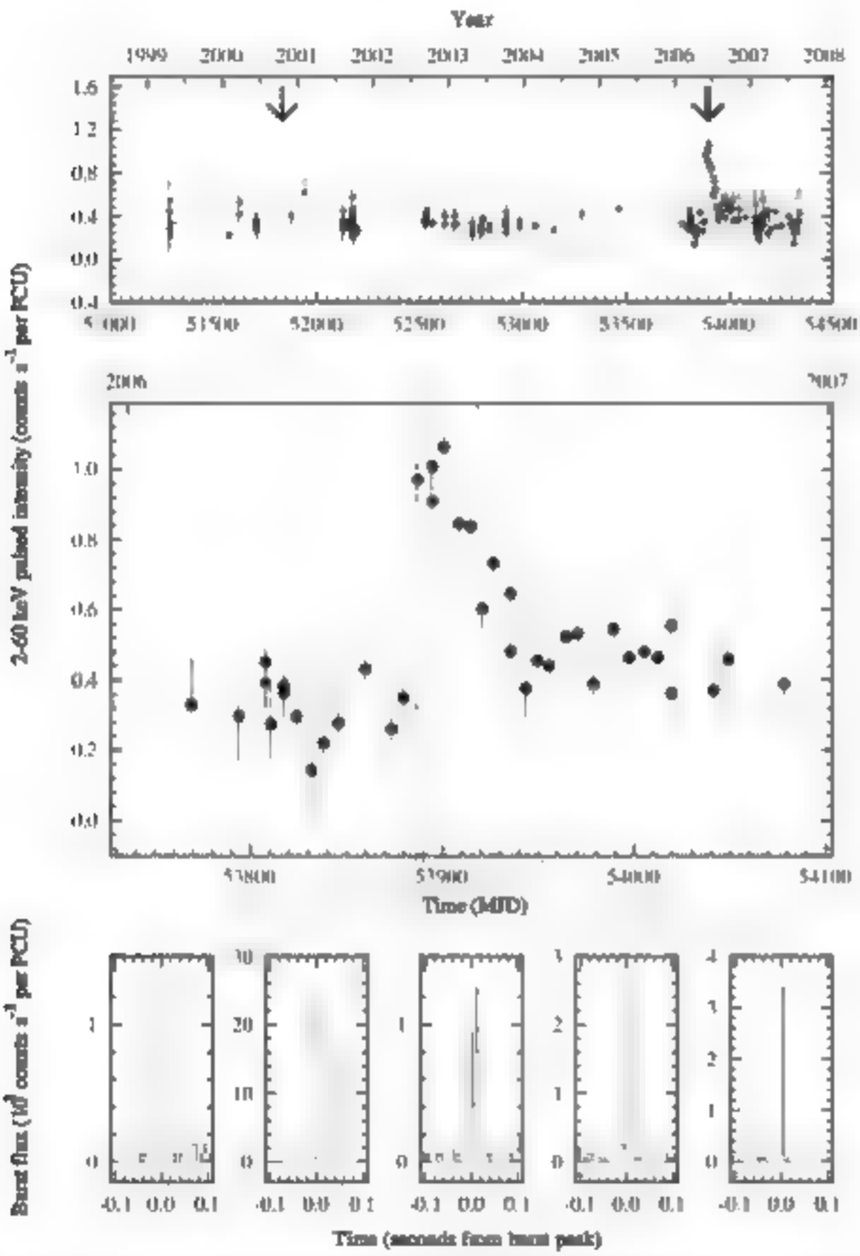


Fig. 2. (Top) Pulsed flux history of PSR J1846–0258 showing the prominent outburst of June 2006 as recorded in the 2- to 60-keV band by RXTE. The horizontal dotted line represents the persistent flux level. Epochs corresponding to CXO observations are indicated with arrows. **(Middle)** The light curve around the outburst. The vertical dashed lines indicate the epochs of the observations containing the bursts, 31 May 2006 (four bursts) and 27 July 2006 (one burst). The leftmost vertical dashed line also coincides with the time when phase coherence was first lost. **(Bottom)** The 2- to 60-keV RXTE x-ray light curves corresponding to five bursts detected from PSR J1846–0258, sampled with 5-ms bins. The bursts lasted for ~0.1 s and were detected with high significance from two data sets obtained on 31 May 2006 and 27 July 2006. In 7 years of RXTE observations, the only bursts found occur either at the onset of the ~2-month x-ray outburst (four bursts) or at the end of the decay (one burst).

Table 1. PSR J1846–0258 burst temporal and spectral properties. All of the quoted errors represent 1σ uncertainties unless otherwise indicated. All times are given in units of universal time corrected, corrected to the Solar System barycenter using the source position right ascension = 18 hours 46 min 24.94 s, declination = –02°58′30″ 1 and the Jet Propulsion Laboratory DE200 ephemeris (*31*). T_{90}, time between when the burst fluence goes from 5 to 95% of its total fluence. Phase 0 means the peak of the average x-ray pulse.

	Burst 1	Burst 2	Burst 3	Burst 4	Burst 5
		Temporal properties			
Burst day (MJD)	53886	53886	53886	53886	53943
Burst start time (fraction of day)	0.92113966(5)	0.93247134(1)	0.93908845(2)	0.94248467(5)	0.45543551(1)
Rise time t_r (ms)	$4.2^{+3.5}_{-2.0}$	$1.1^{+0.9}_{-0.5}$	$1.90^{+1.7}_{-0.9}$	$4.1^{+3.1}_{-1.9}$	$0.9^{+2.2}_{-0.7}$
T_{90} (ms)	$71.8^{+38.0}_{-5.5}$	$42.9^{+0.3}_{-0.2}$	$137.0^{+11.4}_{-36.2}$	$33.4^{+29.1}_{-23.1}$	$65.3^{+0.7}_{-0.5}$
Phase (cycles)	–0.49(1)	–0.04(1)	–0.20(1)	–0.05(1)	–0.08(1)
		Fluences and fluxes			
T_{90} fluence (counts/PCU)	8.9 ± 0.7	712.8 ± 2.5	18.3 ± 0.7	18.4 ± 0.7	18.4 ± 1.1
T_{90} fluence (10^{-10} erg cm^{-2})	4.1 ± 2.4	289.9 ± 13.1	6.6 ± 2.5	5.8 ± 1.7	5.3 ± 2.0
Flux for 64 ms (10^{-10} erg s^{-1} cm^{-2})	57 ± 36	4533 ± 227	99 ± 41	97 ± 31	79 ± 32
Flux for t_r (10^{-10} erg s^{-1} cm^{-2})	678 ± 427	5783 ± 885	810 ± 385	828 ± 284	2698 ± 1193
		Spectral properties			
Power-law index	0.89 ± 0.58	1.05 ± 0.04	1.14 ± 0.34	1.36 ± 0.25	1.41 ± 0.31
χ^2/df (df)	0.42 (1)	1.16 (55)	0.97 (3)	0.35 (2)	1.18 (2)

that released in the 2007 flux enhancement (*18*) of AXP 1E 1048.1–5937 [($\sim 5 \times 10^{42}(d/9\ \mathrm{kpc})^2$ erg], the most energetic enhancement yet seen from this AXP. It is also comparable to the energy released during the rapid [$\sim 3 \times 10^{39}(d/3\ \mathrm{kpc})^2$ erg] and gradual [($\sim 2 \times 10^{41}(d/3\ \mathrm{kpc})^2$ erg] decay components of the 2002 outburst of AXP 1E 2259+586 (*16*). Similar to AXP 1E 2259+586's 2002 outburst (*16*), the energy released by PSR J1846–0258 during the observed short bursts represents only a small (~0.03%) fraction of the total outburst energy.

Before showing magnetar-like emission, PSR J1846–0258 exhibited timing noise and a glitch in 2001 (*6*) that were both similar to what has been seen in other comparably aged (τ ≥ 1000 years) rotation-powered pulsars. By contrast, in 2007, PSR J1846–0258 exhibited much larger timing noise, so that the root mean square phase residual after subtracting a model including the spin frequency, and its first and second derivative, is a factor of ~33 larger than before, for the same duration of observations. Such a dramatic sudden change in timing noise characteristics has never been seen before in a rotation-powered pulsar. The coincidence of the enhanced timing noise with the flux flare is also reminiscent of behavior exhibited by AXP 1E 1048.1–5937 (*15*).

Our discovery of distinctly magnetar-like behavior from what previously seemed like a bona fide rotation-powered pulsar may shed new light on the magnetic evolution of these objects, and on whether their extreme fields originate from a dynamo operating in a rapidly rotating progenitor (*22*), magnetic flux conservation (*23*), or a strongly magnetized core, initially with crustal shielding currents (*24*). In the first two scenarios, magnetars are born with high magnetic fields, which subsequently decay. In the third recently proposed scenario, the very large magnetic fields of magnetars slowly emerge as the shielding currents decay (*24*). This source has a well-measured braking index ($n = 2.65 \pm 0.01$) (*6*), at least before outburst, which is significantly less than 3, suggesting that its spin properties, and hence magnetic field, are headed toward the magnetar regime (*25*). In this case, the time scale for magnetic field evolution, given by the magnetic field divided by its derivative, will be $B/(\dot{B}/B) \sim 8000$ years, at which point PSR J1846–0258 will have $P \sim 1.3$ s. However, other mechanisms, such as the interaction between a strong relativistic pulsar wind nebula and the magnetosphere (*26*), can also yield the value of n measured for PSR J1846–0258. In this case, the magnetar-like behavior could be a result of the moderately high B, with no B evolution occurring.

There have been suggestions of magnetar-related emission from other high-B-field radio pulsars, such as PSR J1119–6127 (*27*), but until now, nothing that could not also be explained within the constraints of rotation-powered pulsar physics. It has been suggested (*4*) that the high-B-field pulsars are related to transient AXPs, which are magnetars generally in quiescence whose x-ray emission can grow by factors of hundreds in outburst. The only two reports of radio pulsations from a magnetar were from transient AXPs after outburst (*28*, *29*). Despite a lack of radio emission (*8*), the behavior of PSR J1846–0258 reinforces the connection between transient AXPs and high-B rotation-powered pulsars, and suggests that careful monitoring of other high-B rotation-powered pulsars (*4*) is warranted.

The addition of PSR J1846–0258 to the list of sources that emit magnetar-like events provides insight into the origin of this activity. Extreme magnetic activity is prevalent in the SGRs, which exhibit giant flares with energy releases upward of 10^{44} erg [see (*30*) for an example] and are also prolific bursters, emitting bursts fairly frequently, typically multiple times per year, with larger outbursts occurring every few years. AXPs can be considered milder versions of SGRs, with several showing sporadic short SGR-like events, though more rarely than SGRs, with even modest outbursts occurring only once or twice per decade. Now Kes 75, weakly magnetized by magnetar standards, shows properties of both rotation-powered pulsars and AXPs and seems to produce an outburst only roughly every decade. The detection of magnetar-like emission from Kes 75 suggests that there is a continuum of magnetar-like activity throughout all neutron stars, which depends on spin-inferred magnetic field strength.

Note added in proof: After submission of this paper, we became aware of two parallel analyses of the Chandra data presented here (*32*, *33*). These provide independent confirmation of our Chandra results.

References and Notes

1. C. Thompson, R. C. Duncan, *Mon. Not. R. Astron. Soc.* **275**, 255 (1995).
2. C. Thompson, R. C. Duncan, *Astrophys. J.* **473**, 322 (1996).
3. P. M. Woods, C. Thompson, in *Compact Stellar X-ray Sources*, W. H. G. Lewin, M. van der Klis, Eds. (Cambridge Univ. Press, Cambridge, MA, 2006), pp. 547–586.
4. V. M. Kaspi, M. A. McLaughlin, *Astrophys. J.* **618**, L41 (2005).
5. E. V. Gotthelf, G. Vasisht, M. Boylan-Kolchin, K. Torii, *Astrophys. J.* **542**, L37 (2000).
6. M. A. Livingstone, V. M. Kaspi, E. V. Gotthelf, L. Kuiper, *Astrophys. J.* **647**, 1286 (2006).
7. D. A. Leahy, W. Tian, *Astron. Astrophys.* **480**, L25 (2008).
8. V. M. Kaspi, R. N. Manchester, S. Johnston, A. G. Lyne, N. D'Amico, *Astrophys. J.* **111**, 2028 (1996).
9. F. P. Gavriil, V. M. Kaspi, P. M. Woods, *Nature* **419**, 142 (2002).
10. F. P. Gavriil, V. M. Kaspi, P. M. Woods, *Astrophys. J.* **607**, 959 (2004).
11. P. M. Woods *et al.*, *Astrophys. J.* **629**, 985 (2005).
12. F. P. Gavriil, V. M. Kaspi, P. M. Woods, *Astrophys. J.* **641**, 418 (2006).
13. V. M. Kaspi *et al.*, *Astrophys. J.* **588**, 93 (2003).
14. A. I. Ibrahim *et al.*, *Astrophys. J.* **609**, L21 (2004).

15. F. P. Gavriil, V. M. Kaspi, *Astrophys. J.* **609**, 287 (2004).
16. P. M. Woods *et al.*, *Astrophys. J.* **605**, 378 (2004).
17. R. Dib, V. M. Kaspi, F. P. Gavriil, *Astrophys. J.* **666**, 1152 (2007).
18. C. K. Tam *et al.*, http://arxiv.org/abs/0712.2856.
19. E. Göğüş *et al.*, *Astrophys. J.* **558**, 228 (2001).
20. C. Thompson, M. Lyutikov, S. R. Kulkarni, *Astrophys. J.* **574**, 332 (2002).
21. A. M. Beloborodov, C. Thompson, *Astrophys. J.* **657**, 967 (2007).
22. C. Thompson, R. C. Duncan, *Astrophys. J.* **408**, 194 (1993).
23. L. Ferrario, D. Wickramasinghe, *Mon. Not. R. Astron. Soc.* **367**, 1323 (2006).
24. D. Bhattacharya, V. Soni, http://arxiv.org/abs/0705.0592.
25. A. G. Lyne, in *Young Neutron Stars and Their Environments, IAU Symposium 218*, F. Camilo, B. M. Gaensler, Eds. (Astronomical Society of the Pacific, San Francisco, 2004), pp. 257–260.
26. A. K. Harding, I. Contopoulos, D. Kazanas, *Astrophys. J.* **525**, L125 (1999).
27. M. E. Gonzalez, V. M. Kaspi, F. Camilo, B. M. Gaensler, M. J. Pivovaroff, *Astrophys. J.* **630**, 489 (2005).
28. F. Camilo *et al.*, *Nature* **442**, 892 (2006).
29. F. Camilo, S. M. Ransom, J. P. Halpern, J. Reynolds, *Astrophys. J.* **666**, L93 (2007).
30. K. Hurley *et al.*, *Nature* **434**, 1098 (2005).
31. D. J. Helfand, B. F. Collins, E. V. Gotthelf, *Astrophys. J.* **582**, 783 (2003).
32. C. Y. Ng, P. O. Slane, B. M. Gaensler, J. P. Hughes, 10th Meeting of the High Energy Astrophysics Division of the American Astronomical Society, Los Angeles, CA, [illegible].
33. H. S. Kumar, S. Safi-Harb, http://arxiv.org/abs/0802.1242.
34. We thank E. Strohmayer for assistance and A. Harding and D. Kazanas for discussion. M.A.I. is a Natural Sciences and Engineering Research Council (NSERC) PGS-D fellow. Support for this work was also provided by an NSERC Discovery Grant, Rgpin 228738-03, an R. Howard Webster Fellowship of the Canadian Institute for Advanced Research, Les Fonds de la Recherche sur la Nature et les Technologies, a Canada Research Chair, and the Lorne Trottier Chair in Astrophysics and Cosmology to V.M.K., and RXTE grants NNG05GM87G and NS-RXTE05-34 to E.V.G. This research made use of data obtained through the High Energy Astrophysics Science Archive Research Center Online Service, provided by the NASA/Goddard Space Flight Center.

Supporting Online Material

www.sciencemag.org/cgi/content/full/1153465/DC1
SOM Text

27 November 2007; accepted 13 February 2008
Published online 21 February 2008;
10.1126/science.1153465
Include this information when citing this paper.

Sr Lattice Clock at 1×10^{-16} Fractional Uncertainty by Remote Optical Evaluation with a Ca Clock

A. D. Ludlow,[1] T. Zelevinsky,[1]* G. K. Campbell,[1] S. Blatt,[1] M. M. Boyd,[1] M. H. G. de Miranda,[1] M. J. Martin,[1] J. W. Thomsen,[1]† S. M. Foreman,[1]‡ Jun Ye,[1]§ T. M. Fortier,[2] J. E. Stalnaker,[2]‖ S. A. Diddams,[2] Y. Le Coq,[2] Z. W. Barber,[2] N. Poli,[2]¶ N. D. Lemke,[2] K. M. Beck,[2] C. W. Oates[2]

Optical atomic clocks promise timekeeping at the highest precision and accuracy, owing to their high operating frequencies. Rigorous evaluations of these clocks require direct comparisons between them. We have realized a high-performance remote comparison of optical clocks over kilometer-scale urban distances, a key step for development, dissemination, and application of these optical standards. Through this remote comparison and a proper design of lattice-confined neutral atoms for clock operation, we evaluate the uncertainty of a strontium (Sr) optical lattice clock at the 1×10^{-16} fractional level, surpassing the current best evaluations of cesium (Cs) primary standards. We also report on the observation of density-dependent effects in the spin-polarized fermionic sample and discuss the current limiting effect of blackbody radiation–induced frequency shifts.

The quest to develop more accurate quantum frequency standards has produced a detailed understanding of matter-field interactions and a vast toolbox of techniques for precision measurement and quantum state control. Historically, the neutral ^{133}Cs microwave clock transition has offered the highest realizable accuracy for state-of-the-art frequency standards (*1*, *2*). In recent years, interest in and performance of standards based on optical atomic transitions have grown (*3*), driven by their superior resonance quality factors (*4*). The most accurate optical clocks are presently based on single trapped ions (*5*), due to the exquisite control possible over their electronic and motional quantum states, as demonstrated in both clock and quantum information experiments (e.g., *5*–*8*). Although neutral atoms (e.g., *9*–*12*) enjoy high measurement precision from the use of large ensembles, a longstanding challenge for these optically based systems is achieving control and measurement at similar uncertainties as for single trapped particles. We report here a systematic uncertainty evaluation for a neutral Sr optical atomic standard at the 10^{-16} fractional level, surpassing the best evaluations of Cs fountain primary standards. This demonstrates control of clock states for large ensembles of atoms approaching that of the best single ion systems. Precise understanding of interactions among lattice-confined atoms will allow clean preparation, control, and readout of atoms for quantum simulations (*13*). Our measurements with thousands of atoms approach the fundamental quantum noise limit, opening the possibility of spin squeezing in optical lattices to combine precision measurement and quantum optics.

Rigorous determination of clock performance can only be achieved by comparing different clocks of similar performance. Although Cs primary standards have served as the best-characterized clock references, direct comparison between optical atomic clocks has now become essential as the uncertainties in the realized, unperturbed clock frequencies are now smaller than those for the best Cs primary standards [as demonstrated in this work and the single ion clock (*5*)]. These systematic uncertainties form the essential element of clock accuracy. Unfortunately, the current complexity of these high-performing optical clocks limits the availability of multiple systems in a single laboratory. Comparing remotely located clocks can circumvent this difficulty. However, traditional methods for these remote clock comparisons such as global positioning satellite links or microwave frequency networks are increasingly inadequate for transferring optical clock signals due to their insufficient stability. All-optical comparison between remote optical clocks permits measurement without compromising the clocks' high precision and ultimately will facilitate tests of fundamental physical laws (*14*) and enable long baseline gravitational measurements (*15*) or long-distance quantum entanglement networks. The optical comparison between the JILA Sr lattice clock on the University of Colorado campus and the calcium (Ca) optical clock at the National Institute of Standards and Technology (NIST) Boulder laboratories is accomplished remotely by a 4-km optical fiber link.

The optical link uses coherent optical carrier transfer [Fig. 1, with more technical details provided in (*16*)]. A self-referenced octave-spanning optical frequency comb at JILA is phase-locked to the Sr clock laser operating at 698 nm. A continuous wave Nd:yttrium-aluminum-garnet (cw Nd:YAG) laser at 1064 nm is phase-locked to the same frequency comb, and its light is transferred to NIST by a phase-noise-cancelled fiber link. In this way, the Sr timekeeping is phase-coherently

[1]JILA, National Institute of Standards and Technology, and University of Colorado, Department of Physics, University of Colorado, Boulder, CO 80309–0440, USA. [2]National Institute of Standards and Technology, 325 Broadway, Boulder, CO 80305, USA.
*Present address: Columbia University, New York, NY 10027, USA.
†Permanent address: Niels Bohr Institute, Copenhagen, Denmark.
‡Present address: Stanford University, Palo Alto, CA 94305, USA.
‖Present address: Oberlin College, Oberlin, OH 44074, USA.
¶Permanent address: Università di Firenze, Italy.
§To whom correspondence should be addressed. E-mail: ye@jila.colorado.edu

transmitted to NIST with a measured transfer instability of $6 \times 10^{-18}/\sqrt{\tau}$ (*17*), where τ is the averaging time. A 10^{-15} fraction of the Sr clock frequency is 0.4 Hz. The transferred Nd:YAG laser light is then frequency-counted against another octave-spanning optical frequency comb at NIST (*18*), which is phase-stabilized to the Ca optical clock operating at 657 nm (*19*). Both comb systems have demonstrated the capability of supporting optical clocks at below the 10^{-18} level (*16*).

For Sr clock operation, ~4000 ^{87}Sr (nuclear spin $I = 9/2$) atoms are laser cooled to 2.5 μK and confined in a one-dimensional (1D) optical lattice. Spectroscopic probing of the $^1S_0$-$^3P_0$ clock transition (Fig. 2A) is performed along the strong confinement axis of the lattice, in the Lamb-Dicke regime and the resolved sideband limit (*8*). Clock interrogation is thus highly immune to Doppler and recoil effects. Using optical pumping, the atoms are (doubly) spin-polarized and occupy only the two states $m_F = \pm 9/2$ (m_F = magnetic quantum number). Figure 2B shows spectra of the π clock transitions ($\Delta m_F = 0$) with and without the optical pumping. With the spin-polarized sample under a bias magnetic field, we interrogate the isolated $m_F = \pm 9/2$ clock transitions for 80 ms, allowing Fourier-limited spectral linewidths of 10 Hz. For these conditions, quantum projection noise would limit the Sr clock stability to $7 \times 10^{-16}/\sqrt{\tau}$ (*20*). To characterize the potential signal-to-noise ratio of the present system, we excite atoms on resonance using a short, strong Rabi pulse to power and Fourier-broaden the excitation, reducing the effect of laser frequency noise. We then measure the normalized, shot-to-shot excitation fraction as we reload the optical lattice with new samples. The excitation fluctuations are consistent with the expected quantum projection noise. However, during clock operation, the probe laser frequency noise deteriorates clock stability. High-frequency laser noise is aliased into the low-frequency measurement band because of the optical Dick effect (*21*). This effect is exacerbated by dead time (1 s) between measurements, which includes cooling the atoms, loading the lattice, polarizing the atomic sample, and determining the populations. The measured frequency noise spectrum of our probing laser (*22*) limits clock stability to $2 \times 10^{-15}/\sqrt{\tau}$.

Figure 2C summarizes these different stability figures for the Sr lattice clock, including the measured Allan deviation between the Sr clock at JILA and the Ca clock at NIST. The Ca clock is a simple and robust system that uses freely expanding cold atoms. Like a hydrogen maser, it serves as a highly stable frequency reference, but with 100 times better stability at short times. Indeed, the stability of the Sr-Ca comparison can reach below 3×10^{-16} after 200 s. However, the Ca clock is susceptible to long-term (>1000 s) drifts due to residual Doppler effects. To optimize evaluation of the Sr clock uncertainties, frequency measurements are thus made in 100-s time windows. To remove sensitivity to long-term Ca drifts, these 100-s windows are interleaved as a particular parameter of the Sr clock is systematically varied. Typically, the parameter is toggled between two settings for two consecutive 100-s windows, and a Sr clock frequency shift is measured between them. Many such pairs are accumulated to average to a measurement precision below 1×10^{-16}, enabling rigorous evaluations of key Sr clock frequency shifts at this level. This measurement approach is facilitated by the robustness of these optical clocks, because both systems are regularly operated on-demand for time scales of a day.

Optical confinement of the Sr atoms occurs at a "magic" wavelength where the polarizabilities of the two clock states are equal (*23*–*25*). The dipole polarizability can be expanded into the scalar, vector, and tensor terms with zero, linear, and quadratic dependence on the m_F of the clock state, respectively (*26*). The opposite symmetries of the vector and tensor polarizabilities facilitate orthogonalization of their effects experimentally. Because of the antisymmetric m_F-dependence of the vector polarizability, clock stabilization to the average of the ±9/2 transitions eliminates dependence of the clock frequency on the vector light shift. This effect would instead be observed as a change in the frequency separation between the two ±9/2 transitions added to the Zeeman splitting from the bias magnetic field. The averaged clock transition retains dependence on the symmetric tensor polarizability, which simply adds a $|m_F|$-dependent offset to the dominant scalar term. Thus, for a given lattice polarization, each spin state has a well-defined magic wavelength for insensitive confinement. The lattice laser providing the atomic confinement is frequency stabilized to the optical frequency comb, which itself is locked to the Sr clock laser. At a lattice peak intensity of $I_0 \approx 3$ kW/cm^2 and at the typical operating laser frequency, we observe a 6.5(5) × 10^{-16} shift in the clock frequency compared to zero intensity (see Fig. 3A). Combined with our previous measurement of the weak clock sensitivity to the lattice frequency (*24*), we extrapolate the magic lattice frequency to be 368,554.68(18) GHz, in agreement with previous measurements (*9*, *23*–*25*). Although tensor shifts are estimated to be small

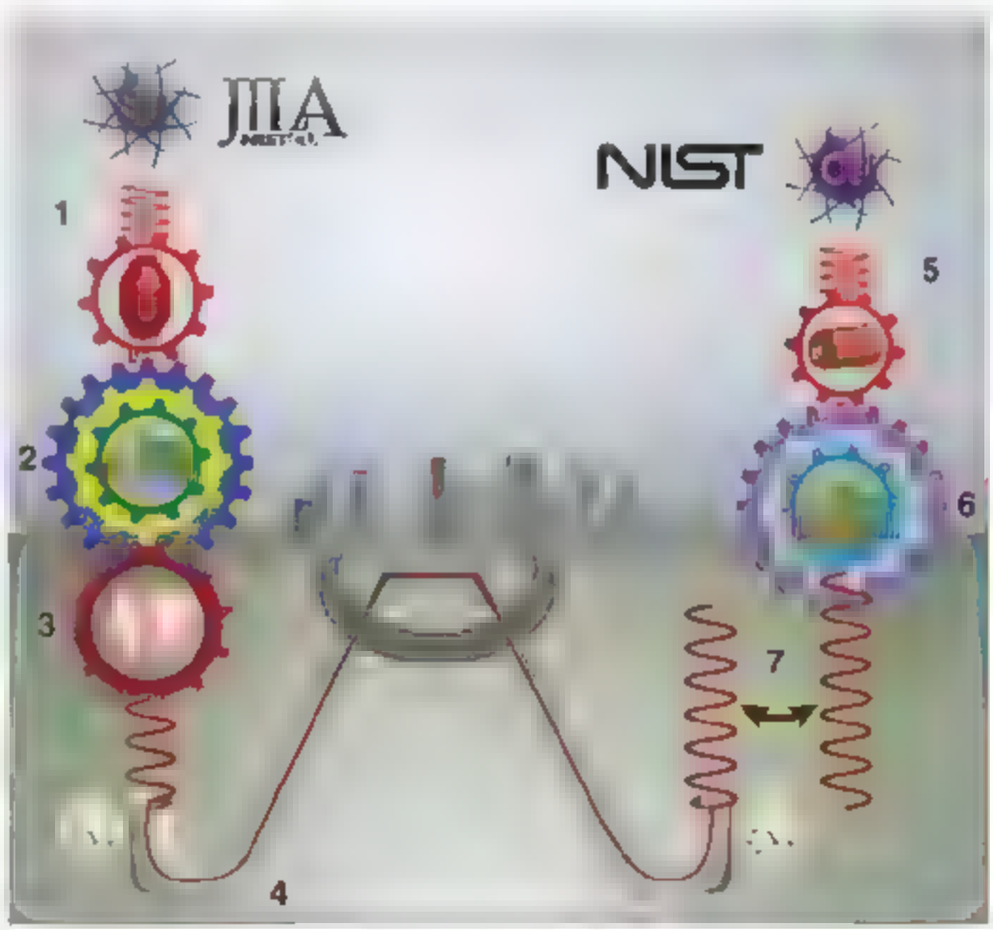


Fig. 1. Remote optical clock comparison. The atomic reference is counted and distributed in direct analogy to traditional clocks using mechanical gears, albeit at optical frequencies. The Sr atomic clock at JILA [1, lattice trapped Sr system and a laser at 698 nm stabilized to vertically mounted high-finesse optical cavity (*27*)] serves as the atomic standard to which an optical frequency comb (2) is phase-locked. A cw Nd:YAG laser at 1064 nm (3, a nonplanar ring oscillator) is phase-locked to the same frequency comb and transferred to NIST by a phase-noise–cancelled fiber link (4). The 1064 light emerges from the fiber at NIST and is phase coherent with the light originating from JILA, as symbolized by the synchronized clocks on either fiber end. The Ca optical clock at NIST (5, free space Ca atoms and a laser stabilized to a horizontally mounted optical cavity operating at 657 nm) is the atomic reference to which an optical frequency comb (6) at NIST is phase-locked. The clock comparison (7) is made by a heterodyne measurement between the NIST frequency comb and the 1064-nm transmitted light.

we clarify that this magic frequency is specified for the π-transitions from the ±9/2 nuclear spin states. The hyperpolarizability effect (fourth order in electric field) has been measured (*25*) and for our operating conditions is $2(2) \times 10^{-17}$.

As with the lattice vector Stark shift, the first-order Zeeman shift is cancelled because of the antisymmetric linear dependence of the ±9/2 states on the magnetic field. To individually address spin states, we choose a bias field (B_0 ≈ 20 μT)

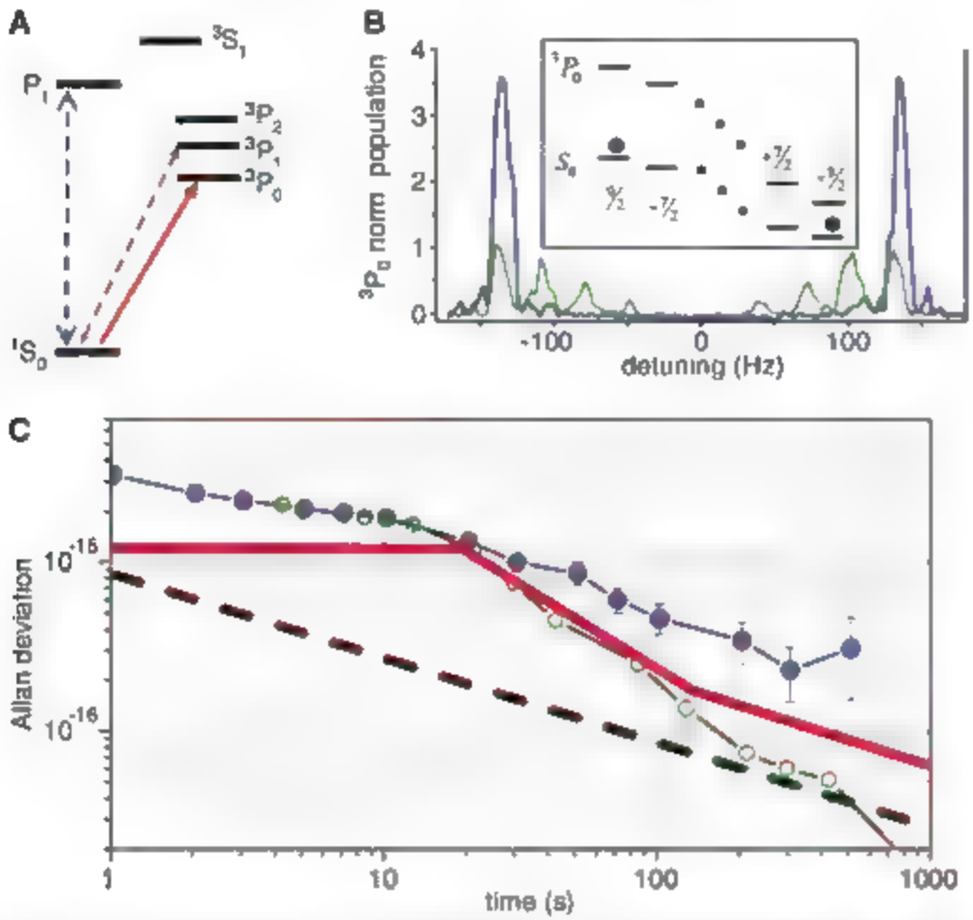


Fig. 2. Sr clock operation. (**A**) Sr energy level diagram. The red and blue dashed lines indicate laser cooling transitions and the solid red line the clock transition. (**B**) Spectroscopic probing of the clock transition under a bias magnetic field to lift the nuclear spin state degeneracy. π transitions with (without) spin polarization optical pumping are shown in blue (green), normalized to the unpolarized ±9/2 peaks. The inset indicates the individual nuclear spin states. After spin polarization, the population resides only in the $m_F = \pm 9/2$ $^1S_0$ states. (**C**) Stability figures for the Sr optical clock as a function of averaging time. Black dashed line: quantum projection noise limit. Blue filled circles: Sr-Ca remote comparison. Green open circles: in-loop stability, based on the error signal of the atom-laser servo. Red solid line: calculated stability, consisting of the free running laser stability (*22*) at short time scales, atom-laser servo operation at intermediate time scales, and the Dick effect limit at long time scales.

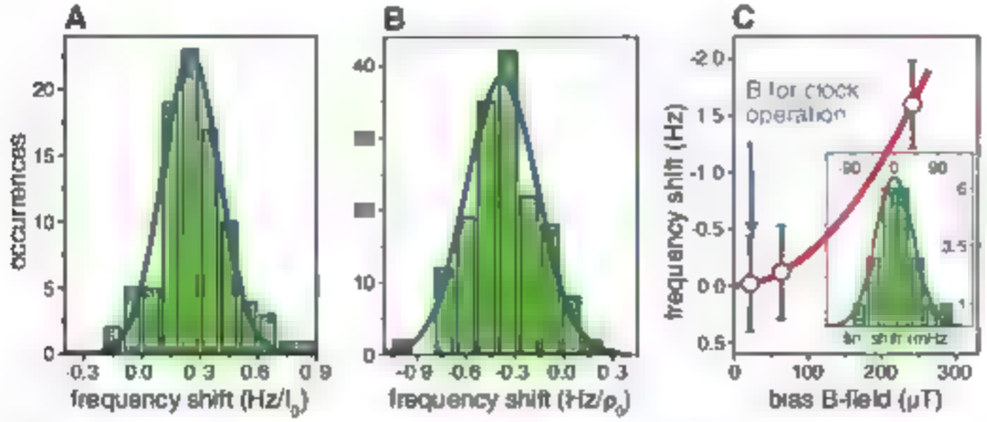


Fig. 3. Clock shift sensitivities. Histogram of clock shift sensitivity to (**A**) the lattice laser intensity slightly away from the magic frequency and (**B**) the atomic density. (**C**) One of 20 data sets indicating the clock sensitivity to magnetic fields. The inset to (**C**) shows the histogram of the linear Zeeman shift evaluated at B_0, based on the 20 data sets.

large enough to resolve π transitions and reduce line pulling from residual populations of other spin states but small enough to keep the spin-symmetric second-order Zeeman shift small. The size of B is calibrated by measuring the frequency spacing between two spin-state transitions (*26*). To determine the second-order sensitivity and the first-order insensitivity, we measure clock frequency shifts as a function of B. An example of one such measurement is shown in Fig. 3C. Each set of data is fit to a second-order polynomial. The fit parameters of many such sets of data are averaged, and the first-order shift is found to be $2(2) \times 10^{-17}$ (for B_0), consistent with zero. The second-order shift is $2.3(2) \times 10^{-17}$, and the measured shift coefficient is $5.8(8) \times 10^{-8}$ T^{-2}, consistent with other measurements (*12*).

Although an ensemble of neutral atoms enables large signal-to-noise measurements for high precision and stability, interactions among colliding atoms can result in frequency shifts that degrade the system accuracy. For the case of lattice clocks, unity (or less) filled sites in a 3D lattice can keep atomic spacing to at least half an optical wavelength and thus reduce interatomic interactions (*23*, *27*). For the 1D lattice, use of identical ^{87}Sr fermions at ultracold temperature can exploit the Pauli exclusion principle to reduce interactions (*28*) by eliminating even-wave collisions. Ground-state ($^1S_0$-$^1S_0$) and excited-state ($^3P_0$-$^3P_0$) interatomic potentials have been theoretically calculated (*29*), but the only experimental measurements exist in photoassociation spectroscopy of even isotopes of Sr (*30*). We have observed a density-dependent frequency shift even when the atoms are polarized to a single spin state, indicating possible P-wave interactions or the loss of indistinguishability due to inhomogeneous excitation (*16*). Varying the atomic density by changing the spin-polarized atom number in a fixed lattice environment allows us to determine this density shift precisely, using the high measurement stability of the optical comparison. This shift scales with the atomic excitation fraction and depends on the nuclear spin state. By controlling the relevant system parameters, we are able to measure and control this shift of $8.9(8) \times 10^{-16}$ (see Fig. 3B, with spin polarization to ±9/2, $\rho_0 = 1 \times 10^{11}$ cm^{-3}). We have also observed an excitation fraction for which the density shift is consistent with zero.

Table 1 gives a summary of our investigation of the Sr systematic clock shifts. The probe laser ac Stark shift, line pulling, and servo error are described in (*16*). The overall fractional systematic uncertainty is 1.5×10^{-16}, the smallest uncertainty reported for any neutral atom standard to date and represents an improvement by a factor of 6 over our previous result (*9*). This uncertainty evaluation has enabled an improved absolute frequency measurement of the Sr clock transition (*31*). The evaluation of the Sr clock is now limited by knowledge of Stark shifts of atomic energy levels induced by the room temperature blackbody radiation (BBR) (*32*), a critical issue for developing standards. The highest

Table 1. Systematic frequency corrections and their associated uncertainties for the $^1S_0$-$^3P_0$ clock transition, in units of 10^{-16} fractional frequency.

Contributor	Correction (10^{-16})	Uncertainty (10^{-16})
Lattice Stark (scalar/tensor)	−6.5	0.5
Hyperpolarizability (lattice)	−0.2	0.2
BBR Stark	52.1	1.0
ac Stark (probe)	0.2	0.1
First-order Zeeman	0.2	0.2
Second-order Zeeman	0.2	0.02
Density	8.9	0.8
Line pulling	0	0.2
Servo error	0	0.5
Second-order Doppler	0	<<0.01
Systematic total	54.9	1.5

accuracy calculation of the BBR shift considers both bound- and continuum-state contributions, dynamical corrections to the static polarizability, and higher-order multipole contributions (*33*). At room temperature, the uncertainty in the BBR shift originating from uncertainty in the polarizability (1%) is 7×10^{-17}. Further uncertainty in the BBR shift originates from lack of control and homogeneity of the blackbody environment (at room temperature, T). By monitoring many positions on the Sr vacuum chamber, we determine the blackbody environment to $\Delta T = 1$ K (root mean square, contributions from the thermal Sr oven are negligibly small), leading to a shift uncertainty of 7.5×10^{-17}. Combining the two effects yields a 1×10^{-16} total BBR uncertainty.

To further improve the Sr accuracy, the differential static polarizability of the clock states must be known to better than 1% (with the dynamical correction contributing <5% to the total shift). This can be measured directly by enclosing the atoms in a well-characterized blackbody environment and recording the clock shift as this temperature is systematically varied (this simultaneously decreases uncertainty in the BBR environment). The technical challenge lies in the control of temperature homogeneity over various functional areas of the vacuum chamber while accommodating sufficient optical access for a variety of atomic manipulations. One possible solution is to cool and trap atoms in a standard chamber and then transport them in a moving lattice (*34*) to a secondary chamber, where an ideal, well-defined blackbody environment is established (*16*). Such an approach avoids the complexity of cryogenic operation and can be generalized to other lattice clock species. These improvements can potentially improve the BBR-related uncertainty to far below 10^{-16}.

References and Notes

1. T. P. Heavner, S. R. Jefferts, E. A. Donley, J. H. Shirley, T. E. Parker, *Metrologia* **42**, 411 (2005).
2. S. Bize *et al.*, *J. Phys. B* **38**, S449 (2005).
3. L. Hollberg *et al.*, *J. Phys. B* **38**, S469 (2005).
4. M. M. Boyd *et al.*, *Science* **314**, 1430 (2006).
5. W. H. Oskay *et al.*, *Phys. Rev. Lett.* **97**, 020801 (2006).
6. H. S. Margolis *et al.*, *Science* **306**, 1355 (2004).
7. T. Schneider, E. Peik, C. Tamm, *Phys. Rev. Lett.* **94**, 230801 (2005).
8. G. Gribbin, R. Blatt, C. Monroe, D. Wineland, *Rev. Mod. Phys.* **75**, 281 (2003).
9. M. M. Boyd *et al.*, *Phys. Rev. Lett.* **98**, 083002 (2007).
10. Z. W. Barber *et al.*, *Phys. Rev. Lett.* **96**, 083002 (2006).
11. M. Takamoto *et al.*, *J. Phys. Soc. Jpn.* **75**, 104302 (2006).
12. X. Baillard *et al.*, *Eur. Phys. J. D*, published online 7 December 2007; 10.1140/epjd/e2007-00330-3.
13. O. Jelsch, P. Zoller, *Ann. Phys.* **315**, 52 (2005).
14. T. M. Fortier *et al.*, *Phys. Rev. Lett.* **98**, 070801 (2007).
15. S. Schiller *et al.*, *Hyp. Phys. B Proc. Suppl.* **166**, 300 (2007).
16. Additional details are available as supporting material on *Science* Online.
17. S. M. Foreman *et al.*, *Phys. Rev. Lett.* **99**, 153601 (2007).
18. T. Fortier, A. Bartels, S. A. Diddams, *Opt. Lett.* **31**, 1011 (2006).
19. C. W. Oates *et al.*, *IEEE Inter. Freq. Control Symposium and Exposition*, 74 (June 2006).
20. W. M. Itano *et al.*, *Phys. Rev. A* **47**, 3554 (1993).
21. G. Santarelli *et al.*, *IEEE Trans. Ultrason. Ferroelec. Freq. Cont.* **45**, 887 (1998).
22. A. D. Ludlow *et al.*, *Opt. Lett.* **32**, 641 (2007).
23. H. Katori, M. Takamoto, V. G. Pal'chikov, V. D. Ovsiannikov, *Phys. Rev. Lett.* **91**, 173005 (2003).
24. A. D. Ludlow *et al.*, *Phys. Rev. Lett.* **96**, 033003 (2006).
25. A. Brusch, R. Le Targat, X. Baillard, M. Fouche, P. Lemonde, *Phys. Rev. Lett.* **96**, 103003 (2006).
26. M. M. Boyd *et al.*, *Phys. Rev. A* **76**, 022510 (2007).
27. D. E. Chang, J. Ye, M. D. Lukin, *Phys. Rev. A* **69**, 023810 (2004).
28. M. W. Zwierlein, Z. Hadzibabic, S. Gupta, W. Ketterle, *Phys. Rev. Lett.* **91**, 250404 (2003).
29. R. Santra, K. V. Christ, C. H. Greene, *Phys. Rev. A* **69**, 042510 (2004).
30. P. G. Mickelson *et al.*, *Phys. Rev. Lett.* **95**, 223002 (2005).
31. The absolute frequency measurement made over a 48-hour period determined the clock frequency to within a total uncertainty of 8.6×10^{-14}, limited by uncertainties in the NIST maser and F1 Cs fountain operation. The final result agrees with our previous measurement (*9*). For more details, see (*35*).
32. L. Hollberg, J. L. Hall, *Phys. Rev. Lett.* **53**, 230 (1984).
33. S. G. Porsev, A. Derevianko, *Phys. Rev. A* **74**, 020502R (2006).
34. S. Schmid, G. Thalhammer, K. Winkler, F. Lang, J. H. Denschlag, *N. J. Phys.* **8**, 159 (2006).
35. S. Blatt *et al.*, *Phys. Rev. Lett.* in press; http://arXiv.org, abs/0801.1874.

36. We thank X. Huang for technical help in operating the Sr clock. This research is supported by the Office of Naval Research, National Institute of Standards and Technology, National Science Foundation, and Defense Advanced Research Projects Agency. A.D.L. acknowledges support from NSF-NGERT through the OSEP program at CU. G.K.C. acknowledges support from the National Research Council.

Supporting Online Material

www.sciencemag.org/cgi/content/full/1153341/DC1
SOM Text
Fig. S1
References

26 November 2007; accepted 23 January 2008
Published online 14 February 2008.
10.1126/science.1153341
Include this information when citing this paper.

Frequency Ratio of Al^+ and Hg^+ Single-Ion Optical Clocks; Metrology at the 17th Decimal Place

T. Rosenband,* D. B. Hume, P. O. Schmidt,† C. W. Chou, A. Brusch, L. Lorini,‡ W. H. Oskay,§ R. E. Drullinger, T. M. Fortier, J. E. Stalnaker,‖ S. A. Diddams, W. C. Swann, N. R. Newbury, W. M. Itano, D. J. Wineland, J. C. Bergquist

Time has always had a special status in physics because of its fundamental role in specifying the regularities of nature and because of the extraordinary precision with which it can be measured. This precision enables tests of fundamental physics and cosmology, as well as practical applications such as satellite navigation. Recently, a regime of operation for atomic clocks based on optical transitions has become possible, promising even higher performance. We report the frequency ratio of two optical atomic clocks with a fractional uncertainty of 5.2×10^{-17}. The ratio of aluminum and mercury single-ion optical clock frequencies ν_{Al^+}/ν_{Hg^+} is 1.052871833148990438(55), where the uncertainty comprises a statistical measurement uncertainty of 4.3×10^{-17}, and systematic uncertainties of 1.9×10^{-17} and 2.3×10^{-17} in the mercury and aluminum frequency standards, respectively. Repeated measurements during the past year yield a preliminary constraint on the temporal variation of the fine-structure constant α of $\dot{\alpha}/\alpha = (-1.6 \pm 2.3) \times 10^{-17}$/year.

Time is the physical coordinate over which humans have the least control, and yet it is the most accurately realized fundamental unit. Although any physical system that evolves predictably can serve as a time base, isolated atoms have long been recognized as near-ideal references for laboratory clocks, due to the abundance of identical copies, as well as their

relative immunity to environmental changes. Because the forces within isolated atoms are many orders of magnitude larger than the external forces perturbing them, atomic resonance frequencies are affected only slightly by external fields. Yet even the small perturbations caused by external fields limit the accuracy of all atomic clocks.

In the work reported here, we combine recent advances in optical and atomic physics to construct atomic clocks based on optical transitions in trapped $^{199}Hg^+$ and $^{27}Al^+$ ions, and measure their frequency ratio. Quantum-jump spectroscopy of single ions (*1*, *2*) and subhertz lasers (*3*, *4*), together with the femtosecond laser frequency comb (*5*, *6*), allowed the first demonstration of an all-optical atomic clock (*7*), which was based on $^{199}Hg^+$. The development of quantum logic spectroscopy (*8*) has enabled the use of $^{27}Al^+$ as a frequency standard (*9*), an ion that is highly immune to external field perturbations (*10*), but whose internal state is difficult to detect by conventional methods.

In each of the standards, the frequency that we attempt to produce in the laboratory is the resonance frequency of the unperturbed ion, at rest and in the absence of background electric and magnetic fields. The deviations from this ideal condition produce shifts that are subtracted from the frequencies of two standards, to the degree that they are known (Table 1). The overall uncertainty in these shifts determines the final accuracy of each frequency standard. Although the specifics of the two standards are quite different, their respective systematic fractional frequency uncertainty is similar: 1.9×10^{-17} for $^{199}Hg^+$ and 2.3×10^{-17} for $^{27}Al^+$. Importantly, none of the current uncertainties are fundamental limits, and both standards can be improved substantially in the future, with a potential accuracy of 10^{-18} or better (*1*, *11*). The ratio of frequencies for the two optical clocks ν_{Al^+}/ν_{Hg^+} reported here marks an order-of-magnitude improvement in achievable measurement accuracy (*12*). As each of these clocks has an accuracy that exceeds current realizations of the SI unit of time, we report the ratio of these optical frequencies, thereby avoiding the uncertainty (3.3×10^{-16}) of the currently realized SI second (*13*).

Until recently, such an optical frequency ratio measurement (Fig. 1) would have required large and costly frequency-multiplication chains to translate between the microwave domain of electronic frequency counters and the optical domain of the clock resonances. The development of tabletop femtosecond laser frequency combs (femtosecond combs) allows this translation to occur in a single, phase-coherent, convenient, and robust step. Here, the fourth harmonics of two clock lasers are locked to the mercury and aluminum clock transitions at 282 and 267 nm, respectively. An octave-spanning self-referenced Ti:Sapphire femtosecond comb (*14*) is phase-locked to one clock laser, and the heterodyne beat-note of the other clock laser with the nearest comb tooth is measured. The various beat-note and offset frequencies are combined to yield the unitless frequency ratio (*12*). In recent comparisons of the frequencies of the two optical clocks described here, a fiber laser femtosecond comb (*15*) has provided a second independent measure of the frequency ratio.

The $^{27}Al^+$ $^1S_0 \leftrightarrow {}^3P_0$ standard, which uses quantum logic spectroscopy (*8*), has been described previously (*9*). One $^{27}Al^+$ ion is trapped together with a $^9Be^+$ ion, which provides sympathetic Doppler laser cooling as well as the means for internal-state detection of the $^{27}Al^+$ ion's clock state ($^1S_0$ or $^3P_0$). The $^{27}Al^+$ clock state is mapped to detectable states in $^9Be^+$ repetitively through the ions' coupled motion, allowing for up to 99.94% clock state detection fidelity (*16*). With the ability to detect the clock state comes the ability to detect state transitions, whose probability depends on the clock laser frequency. The frequency of the clock laser is locked to the atomic transition by alternating between upper and lower slopes of the atomic resonance curve and applying frequency-feedback

Table 1. $^{27}Al^+$ $^1S_0 \leftrightarrow {}^3P_0$ and $^{199}Hg^+$ $^2S_{1/2} \leftrightarrow {}^2D_{5/2}$ clock shifts ($\Delta\nu$), and uncertainties (σ) in units of 10^{-18} of fractional frequency. AOM, Acousto-optic modulator.

Shift	$\Delta\nu_{Al}$	σ_{Al}	$\Delta\nu_{Hg}$	σ_{Hg}	Limitation
Micromotion	−20	20	−4	4	Static electric fields
Secular motion	−16	8	−3	3	Doppler cooling
Blackbody radiation	−12	5	0	0	DC polarizability
313-nm Stark	−7	2	–	–	Polarizability, Intensity
DC quadratic Zeeman	−453	0.5	−1130	5	B-field calibration
AC quadratic Zeeman	0	1	0	10	Trap RF B-fields
Electric quadrupole	0	0.5	0	10	B-field orientation
First-order Doppler	0	1	0	7	Statistical imbalance
Background gas collisions	0	0.5	0	4	Collision model
AOM phase chirp	0	0.1	0	6	RF power
Gravitational red-shift	−5	1	–	–	Clock height
Total	−513	23	−1137	19	

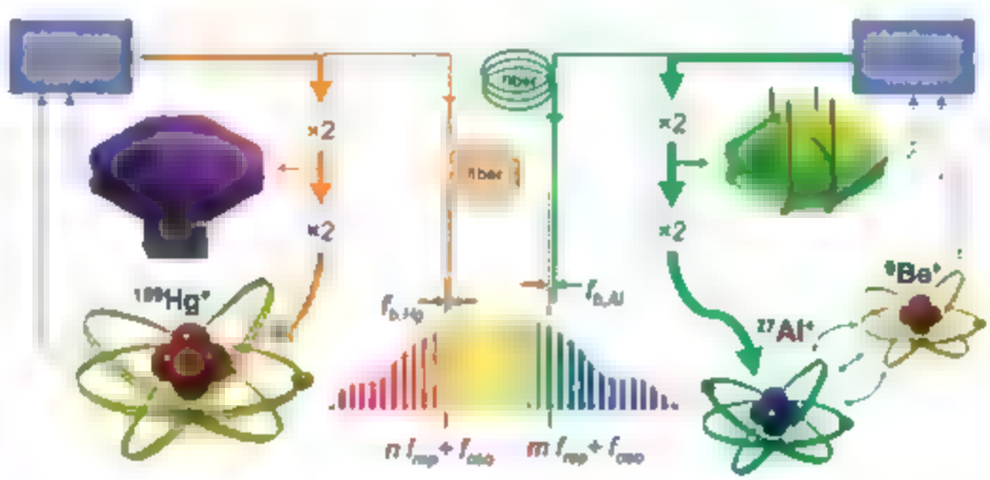


Fig. 1. Frequency ratio measurement system for the comparison of $^{199}Hg^+$ and $^{27}Al^+$ optical clock frequencies. (Left) The fourth harmonic of a 1126-nm wavelength infrared (IR) laser drives atomic-state transitions in a $^{199}Hg^+$ ion (40-ms probe time, 70% duty cycle). The transition rate yields an error signal to keep the laser frequency locked to the atomic resonance. (Right) A 1070-nm wavelength IR laser performs the same function for $^{27}Al^+$ (100-ms probe time, 45% duty cycle), which is coupled to a nearby $^9Be^+$ ion by their mutual Coulomb repulsion for the purposes of sympathetic cooling and internal state detection. Both lasers are prestabilized to ultralow-expansion glass Fabry-Perot cavities (purple and green ellipsoids), thereby narrowing their linewidth to about 1 Hz (*4*). Boxes marked "×2" are second-harmonic generation stages to convert IR light first to visible and then to ultraviolet wavelengths. The two laser frequencies are compared by means of a femtosecond comb (*12*), to which both clock laser systems are linked by 300-m lengths of actively phase-stabilized optical fiber. The quantities $f_{b,Hg}$ (beat note of the mercury clock laser with spectral component n of femtosecond comb), $f_{b,Al}$ (beat note of the aluminum clock laser with spectral component m of the femtosecond comb), f_{ceo} (femtosecond comb carrier-envelope-offset), and f_{rep} (femtosecond comb repetition rate) comprise the frequency ratio measurement (*12*).

National Institute of Standards and Technology, 325 Broadway, Boulder, CO 80305, USA.

*To whom correspondence should be addressed. E-mail: troseri@boulder.nist.gov

†Present address: Institut für Experimentalphysik, Universität Innsbruck, Austria.

‡Present address: Istituto Nazionale di Ricerca Metrologica (INRiM), Torino, Italy.

§Present address: Stanford Research Systems, Sunnyvale, CA 94089, USA.

‖Present address: Department of Physics and Astronomy, Oberlin College, Oberlin, OH 44074, USA.

to keep the transition rates equal. At the operating magnetic field of 0.1 mT, the Zeeman structure due to the nuclear spin of 5/2 is split by several kilohertz, and the individual Zeeman components are well resolved. The clock alternates every 4 s between π-polarized transitions with extreme states of opposite angular momentum ($m_F = \pm 5/2$), which allows compensation of magnetic field shifts to first and second order (*9*, *17*).

The accuracy of the aluminum standard is limited to 2.3×10^{-17} (Table 1), due primarily to uncertainties in the relativistic time dilation, or second-order Doppler shift, caused by microscopic movement of the ion in its trap, with root mean square (RMS) velocities of $v = 1$ to 2 m/s. According to special relativity, moving clocks are observed to run more slowly than stationary ones, with a fractional frequency shift of $-v^2/(2c^2)$. For ions confined in Paul traps, there are two types of motion: secular motion, which is the harmonic motion of the trapped particle, and micromotion, which occurs in part when the ion is displaced from the null of the radiofrequency (RF) confining field by slowly fluctuating electric fields. These quasistatic fields are monitored and nulled by interleaving micromotion test experiments with the clock interrogations. Tests are performed by measuring the strength of radial-to-axial coupling of certain normal modes via $^9Be^+$ (*18*), nulling is accomplished by applying compensation potentials at the ion trap to minimize this coupling. With real-time corrections, the stray electric fields are nulled to (0 ± 10) V/m, allowing an estimate of the time-dilation shift and uncertainty (*19*). The clock shift depends quadratically on the uncompensated electric field, and in the extreme case of a 10-V/m field along both radial directions, the fractional frequency shift is -3.2×10^{-17}. We estimate the shift caused by such residual electric fields to be $(-2 \pm 2) \times 10^{-17}$ when the clock is operating.

Secular mode heating (*20*, *21*) causes deviations in the secular kinetic energy from the Doppler-cooling limit. We apply 313-nm $^9Be^+$ Doppler-cooling light continuously during each 100-ms clock-transition interrogation, to keep the $^{27}Al^+/^9Be^+$ ion pair as cold as possible. However, two poorly damped normal modes of motion heat the ion above the Doppler-cooling limit during this time (*12*), which leads to a second-order Doppler shift of $(-1.6 \pm 0.8) \times 10^{-17}$. Other important shifts are the blackbody radiation shift, which is very small in $^{27}Al^+$ (*10*), and the DC quadratic Zeeman shift, which has been accurately calibrated by varying the magnetic field and measuring the shift in the $^{27}Al^+/^{199}Hg^+$ frequency ratio together with the linear Zeeman splitting ν_1 between the ($^1S_0$ $F = 5/2$, $m_F = +5/2$) → ($^3P_0$ $F = 5/2$, $m_F = \pm 5/2$) lines. The resulting shift is $\nu_2 = -\nu_1^2 \times 1.0479(7) \times 10^{-8}$ Hz^{-1} $= -7.1988(48) \times 10^7$ Hz/T^2.

The $^{199}Hg^+$ ion standard is based on the ($^2S_{1/2}$ $F = 0$) → ($^2D_{5/2}$ $F = 2$, $m_F = 0$) electric-quadrupole transition (*22*). A 194-nm laser cools the ion to the Doppler-cooling limit via the allowed $^2S_{1/2} \rightarrow {}^2P_{1/2}$ transition, and a fiber clock laser frequency-quadrupled to 282 nm excites the clock transition. The clock state of the $^{199}Hg^+$ ion is measured directly via quantum jumps in the scattering fluorescence rate of the 194-nm laser (*2*). Systematic uncertainties in the $^{199}Hg^+$ standard are listed in Table 1 and have been described previously (*22*, *23*). The dominant uncertainties are due to the quadratic Zeeman effect and the electric quadrupole shift. The AC quadratic Zeeman uncertainty stems from possible unbalanced RF currents in the ion trap. The magnitude of this shift is conservatively estimated by assuming a worst-case asymmetry of 50% in the RF currents that flow in the nominally symmetric ion trap. Such an asymmetry would produce an RMS field of $\sim 7.3 \times 10^{-7}$ T at the ion (*12*), causing a fractional frequency shift of -1.0×10^{-17}. We use this value as an upper bound for the magnitude of the shift. The electric-quadrupole shift, which has previously limited the accuracy of the $^{199}Hg^+$ standard, is constrained below 10^{-17} by averaging over three orthogonal magnetic field directions (*23*, *24*).

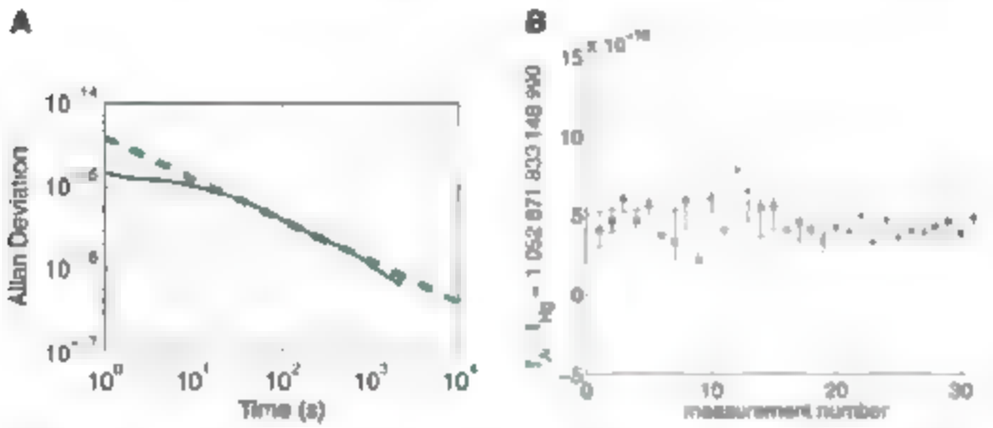


Fig. 2. (**A**) Allan deviation of a frequency comparison measurement (11,000 s total). The dashed line represents a $1/\sqrt{\tau}$ slope, beginning at 3.9×10^{-15} for 1 s. (**B**) History of frequency ratio measurements of the $^{199}Hg^+$ and $^{27}Al^+$ frequency standards. Error bars are statistical. Only the last four points are used in the ratio reported here (Fig. 3).

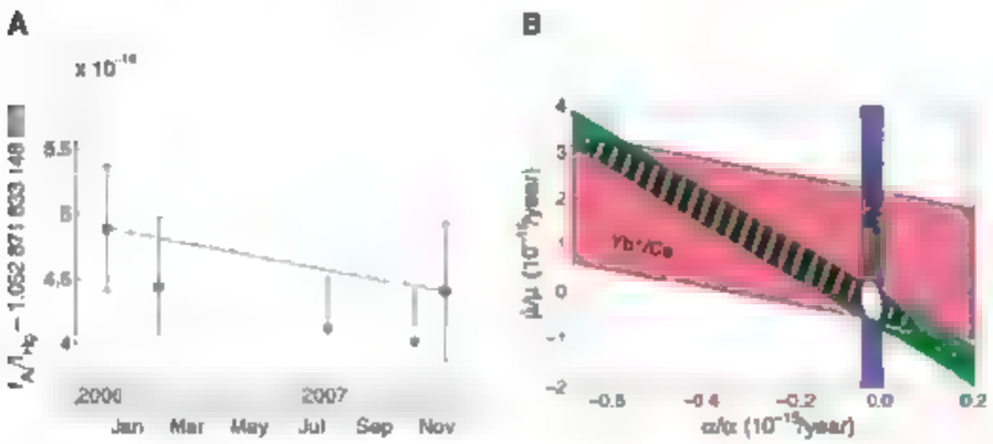


Fig. 3. (**A**) History of frequency ratio measurements grouped by month. The line connects the first point to the last one with a slope of $(-5.3 \pm 7.9) \times 10^{-17}$/year. Lightly shaded points represent measurements where various systematic shifts were not verified to be at the level stated in Table 1 at the time of the measurements. In the first point, the first-order Doppler shift was not monitored, but all other shifts were well controlled. We do not expect this point to have an error due to this effect, because neither apparatus has ever shown a first-order Doppler shift. For the last point, all shifts listed in Table 1 were well controlled. Error bars are a combination of the statistical measurement uncertainty, and the systematic uncertainties listed in Table 1. (**B**) Preliminary constraint on temporal variation of fine-structure constant α from this measurement (vertical bar). The horizontal axis corresponds to variation of the fine-structure constant, and the vertical axis corresponds to variation of the cesium nuclear magnetic moment $\mu = \mu_{Cs}/\mu_B$ in units of the Bohr magneton μ_B, which is not constrained by this measurement. Also shown are laboratory constraints due to $^{199}Hg^+$ versus Cs (*33*) and $^{171}Yb^+$ versus Cs (*34*) measurements. The striped ellipse represents the standard uncertainty for the temporal variation of α and μ due to these previous comparisons of atomic clocks, and the smaller white ellipse shows the reduced uncertainty when the present result is combined with the earlier data. This ellipse constrains $\dot{\mu}/\mu$ to $(-1.9 \pm 4.0) \times 10^{-16}$/year.

Second-order Doppler shifts for $^{199}Hg^+$ are easier to control than for $^{27}Al^+$, because the heavy mercury ion moves less in response to ambient electric fields than the lighter aluminum ion does. Near the Doppler-cooling limit, the total time-dilation shift due to secular motion is $(-3 \pm 3) \times 10^{-18}$. Micromotion is carefully compensated (*19*), leading to a shift of $(-4 \pm 4) \times 10^{-18}$. Thermal blackbody radiation has a negligible effect on the $^{199}Hg^+$ ion, which operates in a cryogenic environment of 4 K (*23*). The quadratic Zeeman coefficient in $^{199}Hg^+$ was calibrated in a way analogous to that of the $^{27}Al^+$ standard. Here the shift is $\nu_2 = -\nu_1^2 \times 4.9465(29) \times 10^{-10}$ Hz^{-1} $\approx -1.89 \times 10^{10}$ Hz/T^2, where ν_1 is the linear Zeeman splitting between the $(^2S_{1/2}\ F = 0) \rightarrow (^2D_{5/2}\ F = 2, m_F = \pm 1)$ lines.

First-order Doppler shifts from photon-recoil are suppressed in tightly bound atomic particles (*25*), due to the Mössbauer effect, where each photon-absorption event lasts for many motional-oscillation cycles. Trapped ions (*1*, *2*) and neutral atoms in optical lattices (*26*) benefit from this effect, but first-order Doppler shifts may still occur, if the trap itself moves in a correlated fashion with the clock laser pulses. Possible causes for such a shift in ion traps might be stray charge buildup from photo-electrons, which are correlated with the interrogation pulses, or more generally, correlated mechanical motion caused by thermal transients or optical shutters. A correlated velocity of only 10^{-8} m/s would cause a first-order Doppler shift error of 3×10^{-17}. Both standards were evaluated for first-order Doppler shifts by illuminating the ions with counter-propagating clock-laser beams, allowing such motion to be both detected and averaged away. The probe direction is selected by shifting the opposite laser beam away from the ion's resonance, either spectrally by ±100 kHz, in the case of $^{27}Al^+$ (several times per second), or spatially by a few beam waists in the $^{199}Hg^+$ clock (every hour). We have not observed a direction-dependent shift in either standard. However, the ratio reported here contains unequal statistical weights for the two probe directions in the $^{199}Hg^+$ standard. This leads to an additional uncertainty of 7×10^{-18} (*12*), which we treat as a systematic uncertainty in the $^{199}Hg^+$ standard (Table 1).

The $^{199}Hg^+$ and $^{27}Al^+$ atomic clocks were operated simultaneously, while the femtosecond combs recorded their frequency ratio every second. Figure 2A shows the Allan deviation (an estimate of the statistical measurement uncertainty versus measurement duration) of a typical ratio measurement. For measurement durations τ greater than 100 s, the deviation is given by $3.9 \times 10^{-15}/\sqrt{\tau/s}$. A departure from this slope at long measurement times, which would indicate fluctuating systematic shifts, has not been observed. We expect that both clocks contribute uncorrelated noise of approximately equal magnitude to the statistical measurement uncertainty and derive a long-term stability of $2.8 \times 10^{-15}/\sqrt{\tau/s}$ for each clock.

Figures 2B and 3A show the history of measurements of the ν_{Al^+}/ν_{Hg^+} ratio. Although the first-order Doppler evaluation discussed in the previous paragraph was carried out fully only for the last four data points of Fig. 2B (which, when combined, give the last point of Fig. 3A), the consistency of the earlier measurements provides confidence in the reproducibility of this result. These four points result in a weighted mean of $\nu_{Al^+}/\nu_{Hg^+} = 1.052871833104990438(55)$, where the statistical uncertainty of 4.3×10^{-17}, the $^{27}Al^+$ systematic uncertainty of 2.3×10^{-17}, and the $^{199}Hg^+$ systematic uncertainty of 1.9×10^{-17} have been added in quadrature, to yield a fractional ratio uncertainty of 5.2×10^{-17}. This ratio may be multiplied by the $^{199}Hg^+$ absolute frequency, which was calibrated by the NIST-F1 primary cesium standard (*27*), to yield a frequency of 1121015393207857.4(7) Hz for the $^{27}Al^+$ standard.

Previous tests for possible counting errors or systematic offsets in the optical frequency combs have shown that the techniques used here support fractional uncertainties on the order of 10^{-19} (*28*, *29*). Further details are provided in the supporting online materials (*12*). As an additional check, for many of the measurements of Fig. 2B, an independent Er fiber-based comb (*15*) measured ν_{Al^+}/ν_{Hg^+} simultaneously with the Ti:sapphire femtosecond comb (*14*). During the last four measurements of Fig. 2B, these measurements show agreement at the level of 1.3×10^{-17}, which is consistent with the nonoverlapping dead-time fractions of 3% and 5% in the frequency counters used by the two independent femtosecond combs. This dead-time error is already included in the statistical measurement uncertainty of 4.3×10^{-17}.

The last entry in Table 1 (the gravitational redshift uncertainty of the $^{27}Al^+$ clock with respect to the $^{199}Hg^+$ clock, is only 10^{-18}, because the two standards are in adjacent laboratories, and their height difference was easily measured with 1-cm uncertainty. This uncertainty is typically much larger for greater separations and, for example, contributes 3×10^{-17} uncertainty for intercontinental primary-standard comparisons between Boulder, Colorado, USA, and other locations (*30*). In the future, high-accuracy portable optical clocks may be available, which could be used to map the height of the geoid with an accuracy beyond that achievable from satellite-based geodesy. Such an endeavor would also require an improved frequency-transfer link between a stationary reference clock and the distant portable standard.

Besides their application to frequency metrology and precision time-keeping, accurate atomic clocks can also help address a fundamental question in physics: Are the constants of nature really constant, or do they change in time or depend on the gravitational potential in which they are measured? Frequency ratio measurements of dissimilar atomic clocks can help answer these questions, because these ratios depend on the fine-structure constant α. From the sequence of measurements in Fig. 3A, one may extract a linear rate of change in the frequency ratio of $(-5.3 \pm 7.9) \times 10^{-17}$/year. A fractional change in α of δ leads to a fractional shift of -3.19δ in the energy of the $^{199}Hg^+$ clock transition when it is expressed in units of the Rydberg energy (*31*). For $^{27}Al^+$, the fractional shift is 0.0085 (*32*). Thus, the measured linear slope in the frequency ratio corresponds to $\dot{\alpha}/\alpha = (-1.6 \pm 2.3) \times 10^{-17}$/year, consistent with no change (Fig. 3B). However, due to the absence of first-order Doppler shift tests in the first point (Fig. 3 legend), this result must be considered preliminary.

The uncertainties in the atomic clocks reported here occur at the exciting intersection of relativity, geodesy, and quantum physics, and the total uncertainty of 5.2×10^{-17} shows unprecedented sensitivity to gravitational effects and cosmological fluctuations. Future improvements in these atomic clocks will provide even more sensitive probes of nature.

References and Notes

1. H. G. Dehmelt, *IEEE Trans. Instrum. Meas.* **31**, 83 (1982).
2. J. C. Bergquist, W. M. Itano, D. J. Wineland, *Phys. Rev. A* **36**, 428 (1987).
3. C. Salomon, D. Hils, J. L. Hall, *J. Opt. Soc. Am. B* **5**, 1576 (1988).
4. B. C. Young, F. C. Cruz, W. M. Itano, J. C. Bergquist, *Phys. Rev. Lett.* **82**, 3799 (1999).
5. T. W. Hänsch, *Rev. Mod. Phys.* **78**, 1297 (2006).
6. J. L. Hall, *Rev. Mod. Phys.* **78**, 1279 (2006).
7. S. A. Diddams *et al.*, *Science* **293**, 825 (2001).
8. P. O. Schmidt *et al.*, *Science* **309**, 749 (2005).
9. T. Rosenband *et al.*, *Phys. Rev. Lett.* **98**, 220801 (2007).
10. T. Rosenband *et al.*, http://xxx.arxiv.org/abs/physics/0611125
11. D. J. Wineland, W. M. Itano, J. C. Bergquist, R. G. Hulet, *Phys. Rev. A* **36**, 2220 (1987).
12. Additional supporting materials are available at *Science* Online.
13. S. R. Jefferts, T. P. Heavner, T. E. Parker, J. H. Shirley, *Proc. SPIE* **6673**, 667309 (2007); available at http://tf.nist.gov/timefreq/general/pdf/2254.pdf
14. T. M. Fortier, A. Bartels, S. A. Diddams, *Opt. Lett.* **31**, 1011 (2006).
15. W. C. Swann *et al.*, *Opt. Lett.* **31**, 3046 (2006).
16. D. B. Hume, T. Rosenband, D. J. Wineland, *Phys. Rev. Lett.* **99**, 120502 (2007).
17. J. E. Bernard, L. Marmet, A. A. Madej, *Opt. Commun.* **150**, 170 (1998).
18. M. D. Barrett *et al.*, *Phys. Rev. A* **68**, 042302 (2003).
19. D. J. Berkeland, J. D. Miller, J. C. Bergquist, W. M. Itano, D. J. Wineland, *J. Appl. Phys.* **83**, 5025 (1998).
20. Q. A. Turchette *et al.*, *Phys. Rev. A* **61**, 063418 (2000).
21. L. Deslauriers *et al.*, *Phys. Rev. Lett.* **97**, 103007 (2006).
22. W. H. Oskay *et al.*, *Phys. Rev. Lett.* **97**, 020801 (2006).
23. W. M. Itano, *J. Res. NIST* **105**, 829 (2000).
24. W. H. Oskay, W. M. Itano, J. C. Bergquist, *Phys. Rev. Lett.* **94**, 163001 (2005).
25. R. H. Dicke, *Phys. Rev.* **89**, 472 (1953).
26. H. Katori, M. Takamoto, V. G. Pal'chikov, V. D. Ovsiannikov, *Phys. Rev. Lett.* **91**, 173005 (2003).
27. J. E. Stalnaker *et al.*, *Appl. Phys. B* **89**, 167 (2007).
28. L.-S. Ma *et al.*, *Science* **303**, 1843 (2004).
29. I. Coddington *et al.*, *Nat. Photonics* **1**, 283 (2007).
30. N. K. Pavlis, M. A. Weiss, *Metrologia* **40**, 66 (2003).
31. V. A. Dzuba, V. V. Flambaum, J. K. Webb, *Phys. Rev. A* **59**, 230 (1999).
32. E. J. Angstmann, V. A. Dzuba, V. V. Flambaum, *Phys. Rev. A* **70**, 014102 (2004).
33. T. M. Fortier *et al.*, *Phys. Rev. Lett.* **98**, 070801 (2007).

34. E. Peik *et al.*, http://arxiv.org/abs/physics/0611058v1.
35. We thank C. W. Oates, M. A. Lombardi, and D. R. Smith for comments on the manuscript and K. Feder, J. Nicholson, and P. Westbrook for specialty nonlinear fiber used in the fiber frequency comb. This work was supported by the Office of Naval Research, Disruptive Technology Office, and NIST. P.O.S. acknowledges support from the Alexander von Humboldt Foundation. This work is a contribution of NIST and is not subject to U.S. copyright.

Supporting Online Material
www.sciencemag.org/cgi/content/full/1154622/DC1
SOM Text
Figs. S1 and S2
References

26 December 2007; accepted 20 February 2008
Published online 6 March 2008;
10.1126/science.1154622
Include this information when citing this paper.

Self-Assembly of Large and Small Molecules into Hierarchically Ordered Sacs and Membranes

Ramille N. Capito,[1] Helena S. Azevedo,[1,2] Yuri S. Velichko,[3] Alvaro Mata,[1] Samuel I. Stupp[1,3,4,5,*]

We report here the self-assembly of macroscopic sacs and membranes at the interface between two aqueous solutions, one containing a megadalton polymer and the other, small self-assembling molecules bearing opposite charge. The resulting structures have a highly ordered architecture in which nanofiber bundles align and reorient by nearly 90° as the membrane grows. The formation of a diffusion barrier upon contact between the two liquids prevents their chaotic mixing. We hypothesize that growth of the membrane is then driven by a dynamic synergy between osmotic pressure of ions and static self-assembly. These robust, self-sealing macroscopic structures offer opportunities in many areas, including the formation of privileged environments for cells, immune barriers, new biological assays, and self-assembly of ordered thick membranes for diverse applications.

The organization of molecules at interfaces has been a phenomenon of great interest over the past few decades given its importance in the preparation of chemically defined surfaces and ordered materials. One classical system is the Langmuir-Blodgett film in which molecules are ordered by compression at an air-water interface, followed by deposition of multilayers through repeated immersion of a solid substrate into the liquid phase (*1*, *2*). Another widely studied system is the self-assembling monolayer formed at a solid-liquid interface by the reaction and self-ordering of dissolved molecules on a solid surface (*3*, *4*). Other systems include the formation of molecular complexes or solid objects at the interface of two immiscible liquids (*6*, *7*), and electrostatic layer-by-layer deposition of amorphous polymers at the solid-liquid interface through alternation of their charge (*8*).

We report the self-assembly of ordered materials well beyond the monolayer scale at an aqueous liquid-liquid interface. We combined a 1 to 2 weight (wt) % peptide amphiphile (PA) solution with a 0.5 to 2 wt% solution of the high molecular weight polysaccharide hyaluronic acid (HA) (fig. S1A). PAs are small synthetic molecules containing typically a hydrophobic alkyl segment covalently grafted to a short peptide. The one used here consisted of an alkyl tail of 16 carbons attached to the peptide sequence $V_3A_3K_2$ (fig. S1B). We previously developed a broad class of PAs that are known to self-assemble into high aspect ratio nanofibers (*9*–*11*). Their tolerance of arbitrary domains past a β-sheet forming sequence makes them useful in biological signaling (*12*, *13*). HA is a linear negatively charged macromolecule containing a disaccharide repeat unit of *N*-acetylglucosamine and glucuronic acid, present in mammalian extracellular matrix. When PA and HA solutions were combined in equal volumes, we observed the immediate formation of a solid membrane localized at the interface between the two liquids. If the denser HA solution is placed on top of the PA solution, the HA fluid sinks into the PA component, causing renewal of the liquid-liquid interface and resulting in continuous growth of membrane until the entire volume of HA solution is engulfed (Fig. 1A). This leads to the formation of a polymer-filled sac over periods of minutes to hours depending on the volume and density of the liquids used. Alternatively, closed sacs can be made instantly by injecting one solution directly into the bulk of the other, creating either HA-filled or PA-filled sacs. Other PAs tested led to similar results without obvious kinetic differences. However, as expected, the nature of the peptide sequence does affect the structural integrity of the membranes formed.

The instant initiation of ordered structures allows formation of self-sealing sacs (Fig. 1, B to D), films with tailorable size and shape (Fig. 1E), as well as continuous strings (Fig. 1F). Confocal microscopy confirmed the incorporation of both HA and PA components within the sac membranes (Fig. 1, G to I). Although formation of solids between oppositely charged macromolecules has been widely demonstrated (*14*–*17*), these systems commonly produce dense and disordered materials that are not permeable to large molecules (e.g., proteins) and are unstable in water without the use of cross-linking chemistry. Moreover, a large defect such as a hole or a crack in a solid made up of two oppositely charged polymers cannot be rapidly repaired by diffusion. The highly ordered materials described here are mechanically robust in both dry and hydrated states, can be permeable to proteins, and have the capability to self-seal defects instantly by self-assembly.

Recognizing the importance of electrostatic charge screening on the self-assembly of PA nanofibers, we investigated the effect of zeta potential in HA and PA solutions on membrane formation (*18*). Sac membranes with physical integrity formed only when both solutions had strong zeta potentials of opposite charge (Fig. 1J). The zeta potential determines the degree of stability of charged aggregates in solution, and in this case should correlate to the total electrostatic charge of molecules in the presence of counterions.

We investigated the microstructure of the PA-HA sac membranes as a function of time using electron microscopy (EM) (*18*). In the early stages of liquid-liquid contact, scanning electron micrographs reveal an amorphous layer (Fig. 2, A and D, region 1) adjacent to a layer of parallel fibers on the PA solution side (Fig. 2, A and D, region 2). We believe that rapid diffusion of the small PA molecules into the HA solution and electrostatic complexation of both molecules result in the formation of the amorphous zone. The parallel fiber region, which measures ~150 nm within 1 min of contact, is the result of self-assembly triggered by electrostatic screening of the amphiphiles by the negatively charged HA molecules near the interface. These early events that occur upon contact between the two liquids establish a diffusion barrier, which prevents the rapid mixing of the two miscible liquids. Ordered growth of nanofibers oriented perpendicular to the interface then forms a layer that measures ~1.5 μm after 30 min (Fig. 2B, region 3) and ~20 μm after 4 days of initial contact (Fig. 2C, region 3).

Transmission electron microscopy (TEM) of cross-sectional slices of a hydrated membrane sample (Fig. 2, E and F) (*18*) confirmed the morphologies observed by scanning electron microscopy (SEM) (Fig. 2D). TEM revealed the presence of the high–molar mass polymer throughout the

[1]Institute for BioNanotechnology in Medicine, Northwestern University, Chicago, IL 60611, USA. [2]3B's Research Group, Biomaterials, Biodegradables and Biomimetics, Department of Polymer Engineering, University of Minho, Braga, Portugal. [3]Department of Materials Science and Engineering, Northwestern University, Evanston, IL 60208, USA. [4]Department of Chemistry, Northwestern University, Evanston, IL 60208, USA. [5]Department of Medicine, Northwestern University, Chicago, IL 60611, USA.

*To whom correspondence should be addressed. E-mail: s-stupp@northwestern.edu

thickness of the membrane (dark regions indicate positive staining of HA by uranyl acetate). The parallel fiber region between the amorphous layer and the perpendicular fibers is more obvious in the TEM micrographs (Fig. 2E, region 2). We calculated the HA density across the membrane shown in Fig. 2E by image analysis (Fig. 2G). The density profile shows three independent regions: (1) a region with an approximately constant polymer density, (2) a region of parallel fibers where there is a maximum in polymer density, and (3) a region of perpendicular fibers where the polymer density decays with increasing distance from the amorphous region. Most of region 1, where there is constant HA density, is probably excess HA from the bulk solution that is merged with a portion of the amorphous region of the membrane. The observed accumulation of HA component in the region of parallel oriented nanofibers (Fig. 2E, region 2) suggests a denser arrangement of nanostructures in that zone. We expect that diffusion of both the large and small molecules will eventually slow down as density increases in the perpendicular fiber region. This is supported by the decrease in HA density across the region of perpendicular fibers.

The mechanical properties of PA-HA membranes were evaluated by a tensile test (*18*). The elastic moduli in the dry and hydrated states of the membrane were found to be about 670 and 0.9 MPa, respectively (Fig. 3A). For comparison, we measured the moduli of membranes formed by the complexation of two oppositely charged polysaccharides, chitosan and gellan (both of high molecular weight). The combination of these two polysaccharides has been previously shown to form macroscopic capsules and strings (*15*, *17*), thus serving as a suitable comparison to our self-assembling systems. The elastic modulus of PA-HA membranes was found to be about nine times higher than that of the chitosan-gellan membranes in both hydrated and dry states (Fig. 3A). Despite their lower stiffness, SEM micrographs of the chitosan-gellan membranes (Fig. 3C) showed a dense structure compared to the PA-HA membrane (Fig. 3B). Furthermore, the chitosan-gellan membranes, swollen to about 20 times their original dry thickness, behave more like hydrogels and fall apart over time when immersed in water. Hydrated PA-HA membranes, in contrast, exhibit minimal dimensional change and remain very stable over months in either water or phosphate-buffered saline. We conclude that hierarchical order contributes to their greater hydrolytic stability and robust mechanical properties.

We found that it is possible to reseal a macroscopic hole in a polymer-filled sac. Using a diacetylene PA molecule (fig. S2) that can be polymerized after self-assembly to an intense blue color (*19*), we were able to create HA-filled sacs in which a macroscopic defect can be easily visualized (Fig. 3D). Because the HA component was contained within the sac, application of additional PA solution within the defect space created a new membrane segment with edges sufficiently sealed to the original sac membrane to prevent leakage of the fluid inside (Fig. 3E). Sealing of the defect does not require polymerization of the PA molecules. Nonetheless, in this case, PA molecules in the repaired region were polymerized to match the appearance of the original sac membrane. Because of their robust character, the PA-HA sacs can also be easily manipulated and sutured (Fig. 3F). When sutured, the sac can hold its own weight without further tearing of the membrane (Fig. 3G).

A common limitation of polyelectrolyte complexes is lack of permeability to larger molecules such as proteins and nucleic acids. We therefore tested the permeability capabilities of the PA-HA membrane by monitoring diffusion of transforming growth factor-β1 (TGF-β1, 25 kD) out of a PA gel-filled sac (*18*). As a control in this experiment, a PA gel was created without the surrounding sac membrane. Sacs enclosing a PA gel and TGF-β1 were created by mixing an ~1 wt% negatively charged PA (fig. S3) and TGF-β1 (2 ng per sac) with the HA solution before injecting the mixture into the positively charged PA. To subsequently gel the negative PA inside, we placed an aliquot of calcium chloride solution on top of the sac, which penetrates the membrane and triggers self-assembly of the negatively charged PA. Results over a 2-week period revealed similar TGF-β1 release kinetics between the gels with and without the surrounding membrane structure (Fig. 3H), thus confirming the permeability of the HA-PA membrane to proteins.

To investigate if these PA-HA structures can support cell viability and differentiation, we performed in vitro studies using human mesenchymal stem cells (hMSCs) (*18*). Expanded hMSCs were incorporated within gel-filled sacs and cultured in growth media, growth media supplemented with TGF-β1, chondrogenic media, or chondrogenic media supplemented with TGF-β1. The hMSCs remained viable within the sacs (Fig. 3I) for up to 4 weeks in culture. Furthermore, results obtained with real-time reverse transcription polymerase chain reaction after 2 weeks in culture (Fig. 3J) revealed an increase in collagen type II gene expression when cells in sacs were stimulated with TGF or chondrogenic media, indi-

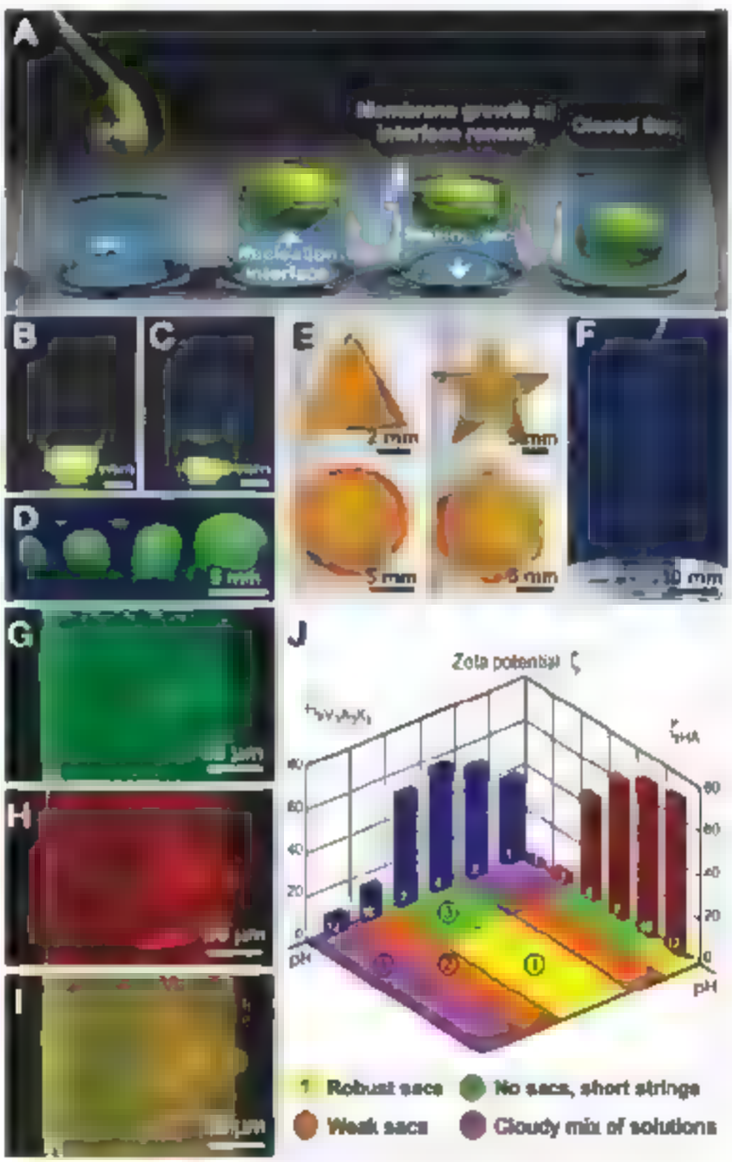


Fig. 1. (**A**) Schematic representation of one method to form a self-sealing closed sac. A sample of the denser negatively charged biopolymer solution is dropped onto a positively charged peptide amphiphile (PA) solution. (**B**) Open and (**C**) closed sac formed by injection of a fluorescently tagged hyaluronic acid (HA) solution into a PA solution. (**D**) Self-assembled sacs of varying sizes. (**E**) PA-HA membranes of different shapes created by interfacing the large- and small-molecule solutions in a very shallow template (~1 mm thick). (**F**) Continuous strings pulled from the interface between the PA and HA solutions. (**G** to **I**) Confocal microscopy of the fluorescein-labeled HA component (**G**) and the rhodamine-labeled PA (**H**) within the sac membrane. (**I**) Merged images demonstrate colocalization of the PA and HA molecules within the membrane. (**J**) Zeta potential measurements of HA and PA solutions measured over a wide range of pH values with a qualitative rating assessing the ability to form well-defined sac structures at each pH condition.

cating that hMSCs were able to differentiate toward a chondrogenic phenotype within the sacs. The self-assembling sacs could therefore provide sufficient nutrient diffusion necessary for cell survival and differentiation.

We have formulated a possible mechanism for the self-assembly of these hierarchically ordered structures. Upon liquid-liquid contact, we first expect rapid mixing by diffusion of molecules localized at the interface (Fig. 4A), with the small molecules diffusing readily into the polymer solution and leading to electrostatic complexation. Ionic screening along the entire contact area leads simultaneously to self-assembly of PA nanofibers localized at the interface (Fig. 4B). This creates a diffusion barrier for long entangled HA chains and PA molecules or their aggregates, resulting in the physical separation of both components. Subsequent growth of fibers oriented perpendicular to the interface occurs toward the PA side as indicated by electron microscopy. This suggests that the megadalton macromolecules diffuse into the small-molecule solution. At first glance this is counterintuitive because one might expect the small molecules to diffuse more rapidly into the polymer solution. Diffusion of the large macromolecules into the PA solution, however, can be driven by the osmotic pressure imbalance between the PA and HA solutions. This imbalance is primarily due to the preservation of macroscopic electroneutrality, because of the greater charge density in a high molecular weight polyelectrolyte, a greater number of ions are required to neutralize the total charge in the HA solution compared to the PA solution. This difference in ion concentration results in an excess osmotic pressure of ions in the HA compartment, resulting in reptation of macromolecules through the diffusion barrier. These macromolecules entering the PA side could then nucleate PA nanofibers perpendicular to the contact interface (Fig. 4C).

By applying Poisson-Boltzman theories to the cell model of polyelectrolytes (*20*, *21*), one can describe the difference in osmotic pressure $\Delta\Pi$ between the two solutions by the expression

$$\Delta\Pi \equiv \Pi_{HA} - \Pi_{PA} = k_B T \sum_i n_i \left(e^{-\frac{\nu_i \psi(R)}{k_B T}} - 1 \right)$$

where n_i and ν_i are the concentration and valence of mobile ions, $\psi(R)$ is a local electrostatic potential, R is the Wigner cell radius, k_B is the Boltzmann constant, and T is the temperature (*20*). The electrostatic potential is proportional to the measured zeta potential, ζ, in our earlier experiments and has a direct effect on the osmotic pressure difference as follows:

$$\Delta\Pi \approx \zeta^2 \sum_i \frac{n_i (\nu_i)^2}{2 k_B T} > 0$$

Therefore, a substantial difference in zeta potential between the two components will result in a large difference in osmotic pressure. Under excess osmotic pressure in the HA compartment, macromolecules will preferentially diffuse through the initially formed contact layer. The required elongated conformation of the HA molecule for diffusion through the pores of the contact layer also creates an entropic barrier that can be overcome only when the excess osmotic pressure dominates. The diffusion driven by the excess osmotic pressure, $\Delta\Pi$, can be described by the Smoluchowski equation

$$\frac{\partial}{\partial t} w(s,t) = D_o \frac{\partial}{\partial x} \left[\frac{\partial}{\partial x} + \frac{\Delta\Pi \lambda b^3}{k_B T} \frac{\partial F}{\partial s} \right] w(s,t)$$

where D_o is the bare diffusion coefficient, λ is the thickness of the diffusive barrier, b is the HA monomer size, $F \propto k_B T \ln(N - s)$ is the entropic free energy of a polymer chain of N repeats with s monomer units that have passed through the membrane, and $w(s,t)$ determines the probability of having a segment of s monomers at time t through the membrane. The released free ends of HA molecules, after complete reptation through the contact layer, produce polymer "stubs" at the surface of the contact layer (Fig. 4D) that attract the oppositely charged PA molecules (Fig. 4B) and nucleate the self-assembly of perpendicularly oriented nanofibers through electrostatic screening (Fig. 4C). Growth of the nanofibers would then occur by attraction of PA aggregates at the base of the polymer stubs as "new" HA segments continue to penetrate through the parallel fiber mesh (Fig. 4, C and E). Self-assembly of the hybrid nanofibers should release condensed counterions that subsequently migrate to the polymer side, given the high affinity of polyelectrolytes for small ions. Hence, the excess osmotic pressure difference can be sustained in a dynamic fashion to continuously pump additional polymer chains through the barrier. One would expect that extensive self-assembly and elongation of these fibers would lead to a crowding effect that further increases the orientational order or align-

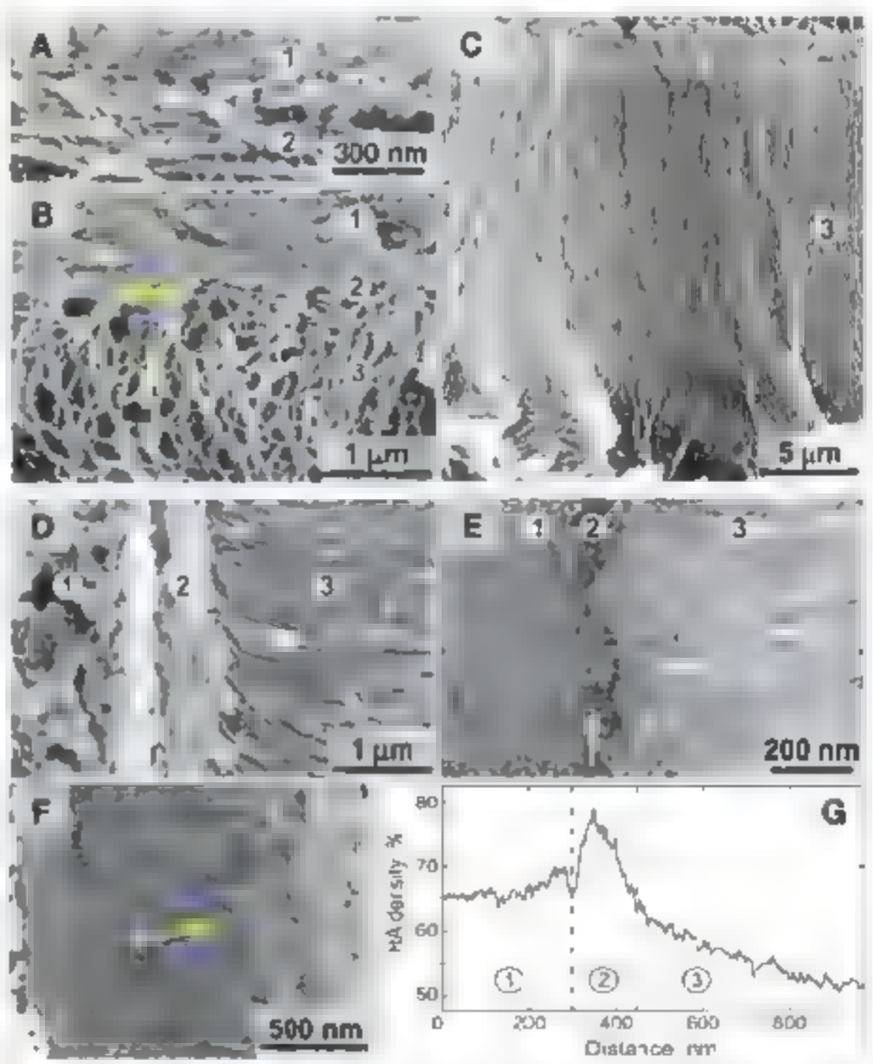


Fig. 2. (**A** to **C**) SEM of the sac membrane as it forms over time (HA solution side on top, PA solution side on bottom): (**A**) 1 min, (**B**) 30 min, and (**C**) 4 days. TEM of a cross-sectional slice of the membrane (**E** and **F**) show morphologies similar to those observed in SEM (**D**), particularly the perpendicularly oriented nanofibers (F, arrows). The parallel fiber region (arrow) between the amorphous and perpendicular fiber zones is most obvious in TEM (E, region 2). Dark regions indicate positive staining of HA by uranyl acetate. (**G**) Polymer density profile obtained by image analysis of (E) quantifying the density of dark regions across the membrane.

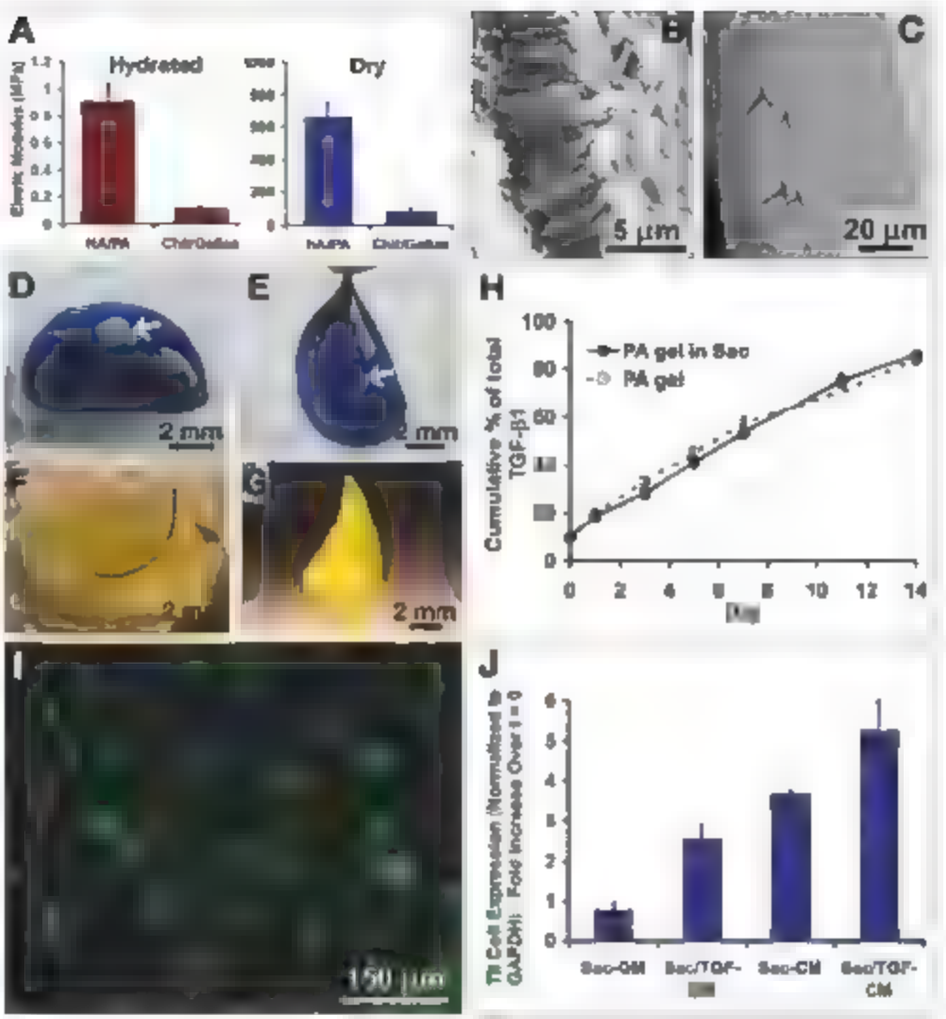


Fig. 3. (**A**) Tensile test data comparing the elastic modulus of the self-assembled membrane to an amorphous membrane composed of chitosan and gellan in both the dry and hydrated states. (**B** and **C**) SEM showing the cross sections of (**B**) the hierarchically ordered membrane and (**C**) the amorphous polymer-polymer membrane [the striations in (**C**) are effects from slicing the dense membrane during sample preparation]. (**D**) Hierarchically ordered sac formed with polydiacetylene PA containing a macroscopic defect within the membrane (arrow). (**E**) Sac in (**D**) after the defect is repaired and the sac resealed by triggering additional self-assembly with a drop of PA (arrow). Sacs are robust enough to withstand suturing (**F**) and can hold their weight without further tearing of the membrane (**G**). (**H**) TGF-β1 release from PA gel–filled sacs and PA gels as a function of time, demonstrating a nearly identical protein-release profile. (**I**) Live/dead assay of hMSCs cultured within the PA gel–filled sacs (green cells are live, red cells are dead) showing that most of the cells remain viable. (**J**) Collagen type II gene expression of hMSCs cultured within the PA gel–filled sacs in growth medium (GM), growth medium supplemented with TGF-β1 (TGF-GM), chondrogenic medium (CM), and chondrogenic medium supplemented with TGF-β1 (TGF-CM). Error bars in (**A**), (**H**), and (**J**) denote SD.

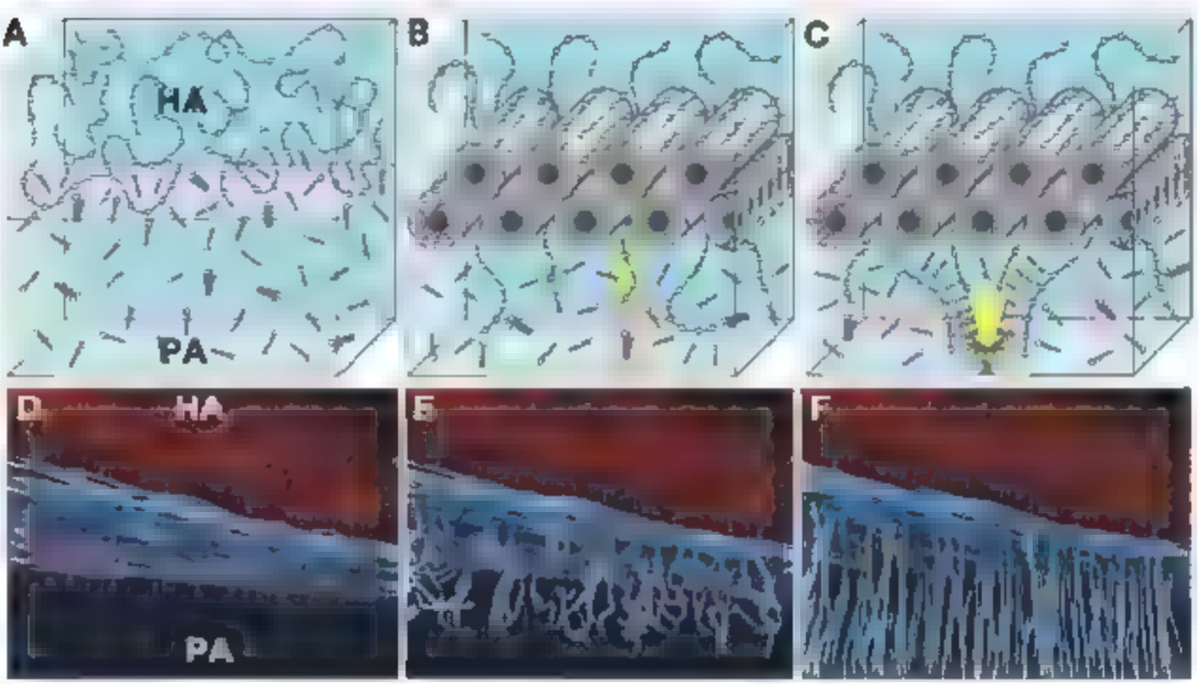


Fig. 4. (**A** to **F**) Schematic illustration of the proposed model for ordered PA·HA membrane formation. (**A**) Initial mixing of the large and small molecules at the interface. (**B**) Formation of nanofibers at the interface composed of small molecules (due to electrostatic screening by the negatively charged polymer) creates a physical barrier between the two solutions. This is followed by reptation of macromolecules through the barrier and into the small-molecule solution (yellow arrow). (**C**) Nucleation of nanofibers perpendicular to the interface by polymer strands crossing the barrier (yellow arrow). Pink arrows indicate attraction of PA monomer or their aggregates to polymer strands as they continue to cross the barrier. (**D**) Schematic representation of polymer stubs (red) penetrating the diffusion barrier. (**E**) Subsequent self-assembly of nanofibers (blue) initiated by the stubs. These nanofibers start gaining orientation perpendicular to the interface as polymer diffuses further into the small-molecule solution. (**F**) Growth of the nanofibers perpendicular to the interface over time, becoming denser and increasingly aligned as the membrane thickens.

ment in this region over time (Fig. 4F), as demonstrated earlier by microscopy (Fig. 2C).

The synergistic interaction between large and self-assembling small molecules with both static and dynamic components has great potential for the discovery of highly functional materials. The sacs formed by these self-assembling systems could be attractive controlled environments for cells in assays or therapies. Moreover, ordered thick membranes could be molecularly customized to possess desired physical or bioactive functions. An interesting possibility is to design similar systems in organic solvents for non-biological applications.

References and Notes

1. K. B. Blodgett, *J. Am. Chem. Soc.* **57**, 1007 (1935).
2. H. Ringsdorf, B. Schlarb, J. Venzmer, *Angew. Chem. Int. Ed. Engl.* **27**, 114 (1988).
3. J. A. Zasadzinski, R. Viswanathan, L. Madsen, J. Garnaes, D. K. Schwartz, *Science* **263**, 1726 (1994).
4. C. D. Bain *et al.*, *J. Am. Chem. Soc.* **111**, 321 (1989).
5. G. M. Whitesides, P. E. Laibinis, *Langmuir* **6**, 87 (1990).
6. L. Prodi *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **99**, 4911 [illegible].
7. N. Bowden, A. Terfort, J. Carbeck, G. M. Whitesides, *Science* **276**, 233 (1997).
8. G. Decher, *Science* **277**, 1232 (1997).
9. J. D. Hartgerink, E. Beniash, S. I. Stupp, *Science* **294**, 1684 (2001).
10. H. A. Behanna, J. J. J. M. Donners, A. C. Gordon, S. I. Stupp, *J. Am. Chem. Soc.* **127**, 1193 (2005).
11. K. Rajangam *et al.*, *Nano Lett.* **6**, 2086 (2006).
12. G. A. Silva *et al.*, *Science* **303**, 1352 (2004).
13. H. Storrie *et al.*, *Biomaterials* **28**, 4608 (2007).
14. S. Wen, Y. Xiaonan, W. T. K. Stevenson, *Biotechnol. Tech.* **12**, 374 (1998).
15. H. Yamamoto, Y. Senoo, *Macromol. Chem. Phys.* **201**, 84 (2000).
16. H. Yamamoto, C. Horita, Y. Senoo, A. Nishida, K. Ohkawa, *J. Appl. Polym. Sci.* **79**, 437 (2001).
17. K. Ohkawa, T. Kitagawa, H. Yamamoto, *Macromol. Mater. Eng.* **289**, 33 (2004).
18. Materials and methods are available as supporting material on *Science* Online.
19. L. Hsu, G. L. Cvetanovich, S. I. Stupp, *J. Am. Chem. Soc.* 10.1021/ja076553s (2008).
20. M. Deserno, H. H. von Grunberg, *Phys. Rev. E Stat. Nonlin. Soft Matter Phys.* **66**, 011401 (2002).
21. C. Holm, P. Kekicheff, R. Podgornik, *Electrostatic Effects in Soft Matter and Biophysics* (NATO Science Series II, Kluwer, Dordrecht, Netherlands, 2001), pp. 27–52.

22. Supported by the U.S. Department of Energy (grant DE-FG02-00ER45810), the NIH (grants 5-RO1-EB003806, 5-RO1-DE015920, and 5-PSO-N50S4287), and the NSF (grant DMR-0605427). We thank N. A. Shah for the suturing tests, L. Hsu for synthesis of the diacetylene peptide amphiphile, R. Walsh for processing TEM samples, the Brinson laboratory at Northwestern University for assistance with mechanical testing, and M. Senow for assistance with illustrations.

Supporting Online Material

www.sciencemag.org/cgi/content/full/319/5871/1812/DC1
Materials and Methods
Figs. S1 to S3

26 December 2007; accepted 7 February 2008
10.1126/science.1154566

The Transition from Stiff to Compliant Materials in Squid Beaks

Ali Miserez,[1,2,3] Todd Schneberk,[2,3] Chengjun Sun,[2,3] Frank W. Zok,[1]* J. Herbert Waite[2,3]*

The beak of the Humboldt squid *Dosidicus gigas* represents one of the hardest and stiffest wholly organic materials known. As it is deeply embedded within the soft buccal envelope, the manner in which impact forces are transmitted between beak and envelope is a matter of considerable scientific interest. Here, we show that the hydrated beak exhibits a large stiffness gradient, spanning two orders of magnitude from the tip to the base. This gradient is correlated with a chemical gradient involving mixtures of chitin, water, and His-rich proteins that contain 3,4-dihydroxyphenyl-L-alanine (dopa) and undergo extensive stabilization by histidyl-dopa cross-link formation. These findings may serve as a foundation for identifying design principles for attaching mechanically mismatched materials in engineering and biological applications.

Living organisms are functional assemblages of different interconnected tissues. Not infrequently, tissues with highly disparate mechanical properties (e.g., bone and cartilage, shell and adductor muscle, nail and skin) are joined together (*1*). In practice, the joining of dissimilar materials can lead to high interfacial stresses and contact damage (*2*, *3*). In apparent contradiction to this, the contacts between mechanically mismatched biomolecular tissues are remarkably robust. Mechanical-property gradients are increasingly invoked as the principal reason for their mechanical performance. The dentino-enamel junction (*4*), arthropod exoskeleton (*5*), polychaete jaws, and mussel byssal threads (*6*) all exhibit such gradients. Optical properties in squid eyes have also been correlated to a protein-density gradient (*7*). Although much is known about the mechanical and biochemical properties of the separate tissues, surprisingly little has been done to explain how mixtures of macromolecules are adapted for incremental mechanical effects at interfaces.

The beak of the Humboldt squid *Dosidicus gigas* is an example of a system with grossly mismatched tissues. It is composed of slightly offset apposing upper and lower parts that make no hard pivotal contact with one another and are set into a muscular buccal mass that controls their movement (*8*). The beak tip (or rostrum) is among the hardest and stiffest of wholly organic biomaterials (*9*). With a single notch through the thick dorsal integument of a captured fish, a squid beak can sever the nerve cord to paralyze prey for later leisurely dining (*10*). The predatory activity of a stiff razor-sharp beak embedded in a softer muscle mass might be compared to a person carving a roast with a bare metal blade. In both scenarios, there would be as much damage to self as to the intended objects. We found that the squid's task is facilitated by a beak design that incorporates large gradients in mechanical properties, intricately linked with local macromolecular composition, from the hard, stiff tip to the soft, compliant base.

When detached from the buccal mass, each Humboldt squid beak exhibits clearly visible gradients in pigmentation, ranging from translucent along the wing edge to black at the beak tip (Fig. 1, B and C). The pigmentation gradient appears to be closely coupled to a distribution of catechols that correspond to proteins containing 3,4-dihydroxyphenyl-L-alanine (dopa) (*9*). Catechol staining (red in Fig. 1D) is evident in the lightly tanned regions and more intensely (though largely obscured by black pigment) in the heavily tanned regions. Because the beaks are large enough to allow cutout specimens showing incremental differences in pigmentation (Fig. 1E), we performed coupled chemical and mechanical analyses to explore the consequences of pigmentation (figs. S1 and S2).

We carried out chemical analyses of consecutive cutouts after degradation by two treatments that preferentially attack different structural components: acid hydrolysis, which hydrolyzes protein and chitin but not the tanning pigment, and alkaline peroxidation, which solubilizes everything but chitin (see flowchart in fig. S1). The hydrolyzed untanned material is dominated by glucosamine, the basic structural unit of chitin [Fig. 2A(i)], whose presence was independently established by fiber x-ray diffraction (*9*). On the basis of glucosamine detection, chitin content in this region is estimated to be 90% of dry weight, compared to a low of 15 to 20% in the rostrum [Fig. 2A(ii)]. The amino acid composition of hydrolysates from all tanned regions is uniform and dominated by Gly, Ala, His and Asx (where Asx represents Asp and Asn combined) (27, 15, 11, and 7 mole percent, respectively) (Fig. 2, A and B). Only the untanned region differed in composition, with the Asx content being considerably higher and that of the other amino acids being somewhat lower than the corresponding values in the tanned regions. This disparity may

[1]Materials Department, University of California, Santa Barbara, CA 93106, USA. [2]Department of Molecular, Cellular, and Developmental Biology, University of California, Santa Barbara, CA 93106, USA. [3]Marine Science Institute, University of California, Santa Barbara, CA 93106, USA.

*To whom correspondence should be addressed. E-mail: zok@engineering.ucsb.edu (F.W.Z.); waite@lifesci.ucsb.edu (J.H.W.)

be a result of the low protein content of the untanned material, as well as the difficulty of removing all adhering connective tissue derived from the buccal mass.

Alkaline peroxidation removes all proteins and pigments, leaving essentially pure chitin (95% of dry weight) [Fig. 2A(iii)]. The chitin is present in the form of a scaffold of fine (30-nm diameter) interconnected fibers that maintains beak shape (Fig. 2C). The scaffold structure appears to be the same from the base to the beak tip. The resistance of chitin to peroxide degradation was confirmed by treatment of a commercial chitin control. The protein liberated by alkaline peroxidation has essentially the same amino acid composition [Fig. 2A(iv)] as that detected in cutout specimens after direct beak hydrolysis [Fig. 2A(ii)], suggesting that protein is bound to and released with the tanning pigments during peroxidation.

The relative contributions of chitin, protein, and water in hydrated beak specimens were ascertained through gravimetric analysis, by weighing hydrated and freeze-dried samples before and after alkaline peroxidation. Tanning pigments were derivatized into soluble orange chromophores by alkaline peroxidation, and their concentrations were estimated by calculating the extinction of the orange chromophore with a known amount of nonhydrolyzable black pigment (fig. S4). The resulting compositions are summarized in Fig 2D. The wing contains about 70 weight percent (wt %) water, 25 wt % chitin, and 5 wt % protein. The water content decreases with increasing concentrations of both protein and tanned pigment. water content reaches a minimum of 15 to 20 wt % in the rostrum, where the protein and tanning pigment levels are 60 and 20%, respectively. In contrast to protein and tanned pigment, chitin content follows the opposite trend, decreasing to less than 15 wt % in the rostrum, but with a strong positive correlation to water content

Because catechols are often coupled to pigmentation in living organisms, the relation between dopa-proteins and tanning pigment in squid beaks merited further scrutiny. Both dopa and several cross-link derivatives were captured from beak cutouts by phenylboronate chromatography after protein hydrolysis. The captured compounds were characterized by ultraviolet/visible spectroscopy (fig. S2), amino acid analysis (fig. S3), electrospray ionization mass spectrometry (ESI-MS), and tandem mass spectrometry (MS/MS). The major MS peaks detected after a 24 hour hydrolysis had mass/charge ratios (m/z) of 278 and 351; a less stable compound at m/z 335 was captured after a shorter (2 hour) hydrolysis. Structures deduced for each are consistent with a His-dopa peptide (m/z 335) and a His-dopa cross-link (m/z 351) plus a related derivative (m/z 278) (Fig. 3A and fig. S5). The tendency of these to undergo oxidation to black pigments in vitro at pH 7.0 suggests that the cross-links are soluble antecedents of the acid-insoluble black pigment. This hypothesis was tested by subjecting the acid-insoluble black pigment to direct laser desorption mass spectrometry with the pigment acting as its own matrix (Fig. 3B). Only a few prominent fragments, particularly at m/z 198 and 156, were observed. These resemble the dopa^{1H+} and His^{1H+} ions released during MS/MS of the dopa-His cross-links and are distinct from those in authentic eumelanin (m/z of 552, 365, and 208) (*11*). It is thus reasonable to assume that insoluble black pigments from squid beaks represent multimers of dopa-His cross-links.

Both tensile and nanoindentation tests show that when the coupons are hydrated, the Young's modulus E increases dramatically with the degree of pigmentation, from about 0.05 to 5 GPa (Fig. 4, A and B). However, after the peroxide treatment, the values drop drastically, to a mere 0.03 GPa, regardless of pigmentation. Evidently the chitin scaffold on its own has minimal structural integrity. In the freeze-dried state, E is uniformly high, diminishing by only a factor of 2 (from about 10 to 5 GPa) between the rostrum and the untanned wing (Fig. 4B). Essentially the same reduction was obtained in the rostrum after peroxide treatment, suggesting that the proteins play a minimal role in beak reinforcement in the absence of water.

To glean additional insights into compositional effects, E was replotted against both chitin content and combined chitin/protein content (Fig. 4C). In the freeze-dried state, the modulus depends only weakly on protein or chitin content. In the hydrated states, however, it exhibits a strong dependence on composition. Most notably, it appears to follow an inverse relation with chitin content—an unexpected result because chitin fibers are among the stiffest polysaccharides (E ~ 40 GPa in the dry state) (*12*). Concurrently, though, the water content increases more than threefold. The corresponding degree of chitin hydration, characterized by the water-to-chitin mass ratio, increases from about 2.2 in the rostrum to 3.3 in the untanned regions. Evidently, over this range, changes in the degree of hydration play a greater role than that of chitin content, leading to the observed stiffness reduction with increased distance from the beak tip

Beak stiffness is undoubtedly influenced by both microarchitectural and molecular factors. With respect to the latter, stiffness is closely correlated to the incremental and complementary distributions of two biopolymers—chitin, a polysaccharide of β1,4-linked *N*-acetylglucosamine, and a family of His- and dopa-containing proteins and the extent to which they are hydrated or cross-linked. Although the proteins have yet to be fully characterized, the dipeptide His-dopa and His-dopa cross-links (Fig. 3A and fig. S5) attest to the frequent proximity of the two amino acids in squid beaks. In this respect, squid beaks are reminiscent of sclerotized insect cuticle. Both contain catechols (peptidyl-dopa and N-acetyldopamine

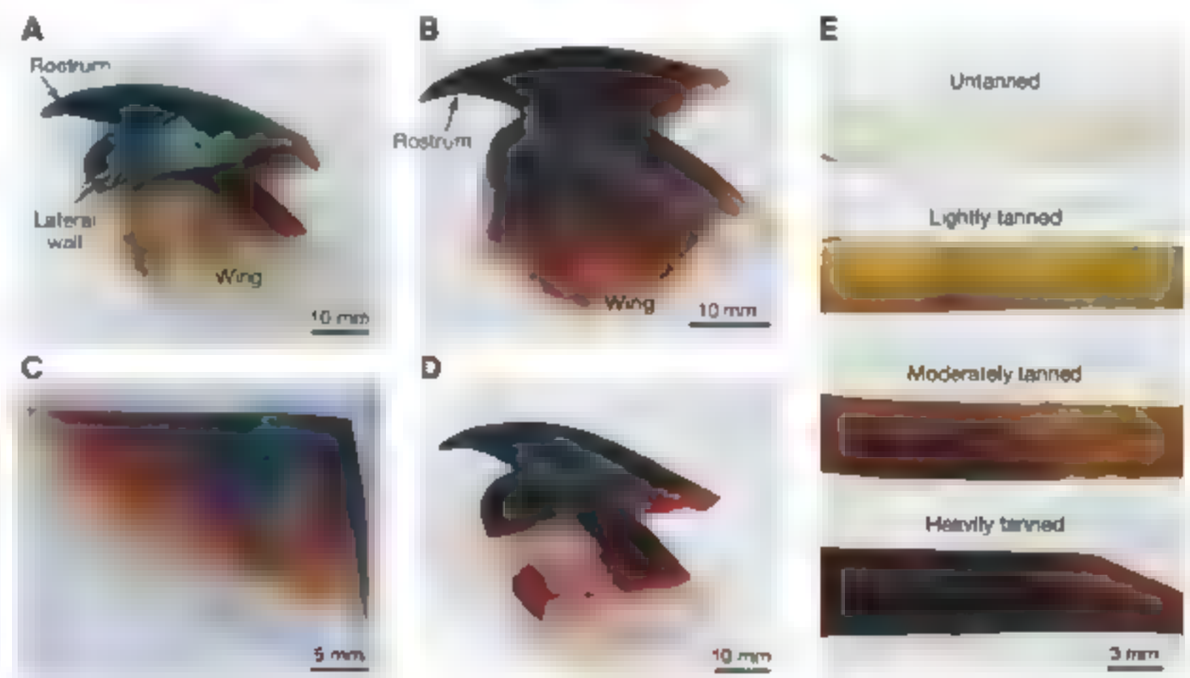


Fig. 1. Beak of the Humboldt squid *Dosidicus gigas*. **(A)** Side view of a full upper beak after removal from the buccal mass showing natural pigmentation. **(B)** Split beak highlighting the relation of the wing to the rostrum. **(C)** Dissected wing. **(D)** Beak in (A) after staining for dopa-containing proteins with a catechol-specific reagent. **(E)** Series of sections representing different shades of the pigmentation gradient.

(NADA) (*13*), proteins with His-rich domains (*14*), catechol-His cross-links (*15*, *16*) and a reinforcing chitin network (*17*, *18*). However, there are also substantial differences. In contrast to the low–molecular weight NADA used by insects, the catecholic functionality in beaks (dopa) is tethered to a protein backbone, imposing steric and diffusional constraints on compositional gradients. With three distinct ingredients (NADA, chitin, and protein), insects can and do vary NADA independently of the other two ingredients (*19*). This process is severely limited with protein-tethered dopa. In a protein with a fixed proportion of its tyrosyl residues modified to dopa, a gradient of dopa necessarily includes a gradient of protein and vice versa. There are cases in which the Tyr residues in a protein are differentially modified to dopa over the range of its expression in a tissue. For example, mussels have an increasing gradient of dopa/Tyr in *Mytilus* foot protein-1 along the proximal-to-distal axis of its distribution in each byssal thread (*20*). In squid beaks, the dopa/Tyr ratio in precursor proteins is not yet known, however, the pigment/protein ratio remains roughly constant.

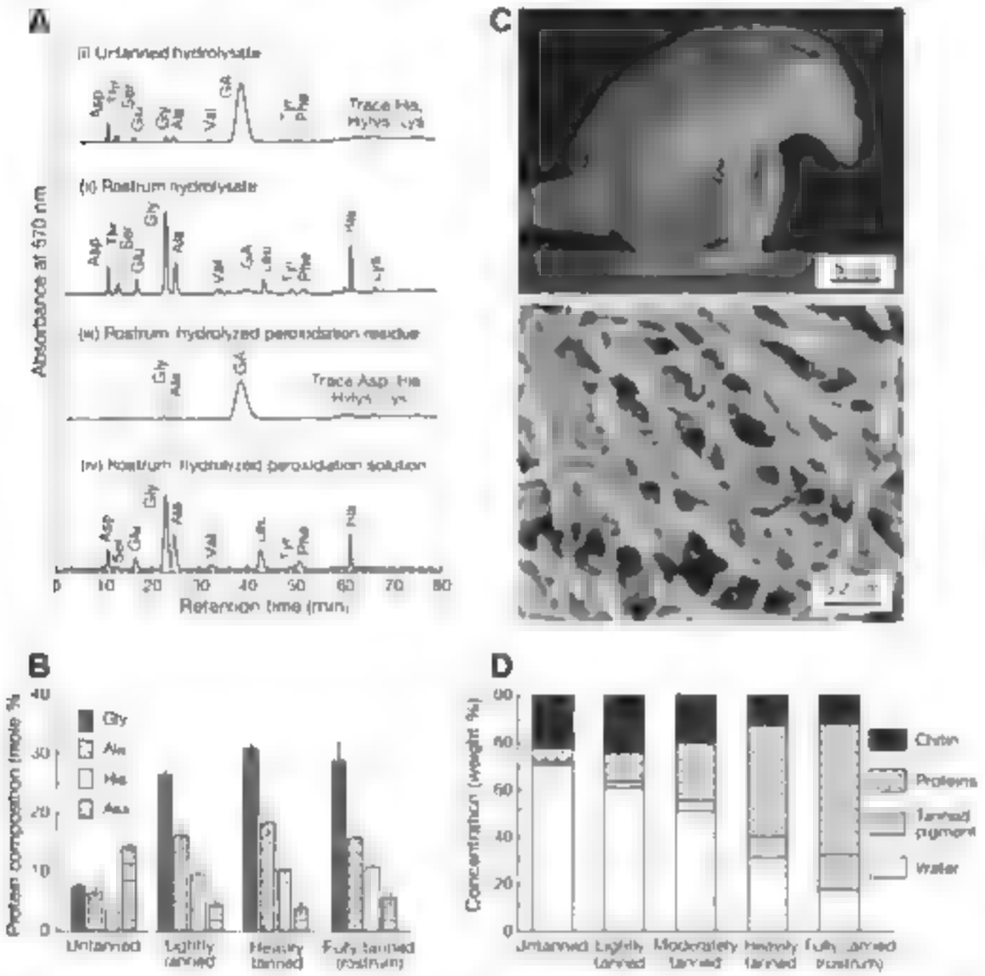


Fig. 2. Chemical and gravimetric analyses of squid beak cutouts. (**A**) Absorbance spectra of acid hydrolysates. GA, glucosamine; Hylys, hydroxylysine. (**B**) Compositions of the dominant amino acids in the hydrolysates. (**C**) Optical image of the whole beak after alkaline peroxidation (top), and high-magnification scanning electron image of the chitin fiber network in the rostrum (bottom). (**D**) Chitin, protein, and pigment contents ascertained from gravimetric measurements.

The most notable discovery of this study is the importance of water content in defining the stiffness gradient. Complete desiccation by freeze-drying virtually eliminates the gradient, narrowing stiffness to between 5 and 10 GPa. On the basis of the inverse distributions of chitin and protein, the stepwise desolvation of chitin may be accomplished by a titrated admixture of dopa-containing beak proteins. Catechols are known to have a potent dehydrating effect on various biopolymers, attributable to their hydrophobicity (*21*, *22*) and/or hydrogen-bonding ability (*23*). However, the occurrence of cross-links, such as dopa-His cross-links, and the incremental deposition of cross-linked multimers that constitute the tanned pigment are also likely to reduce solvation (*24*, *25*). Absent from our squid beak analyses are cross-links between protein and chitin; whether this means that the two macromolecules are not covalently cross-linked or that the cross-links are not captured by phenylboronate remains to be determined. Although both noncovalent and covalent factors are important for controlling the hydration of squid

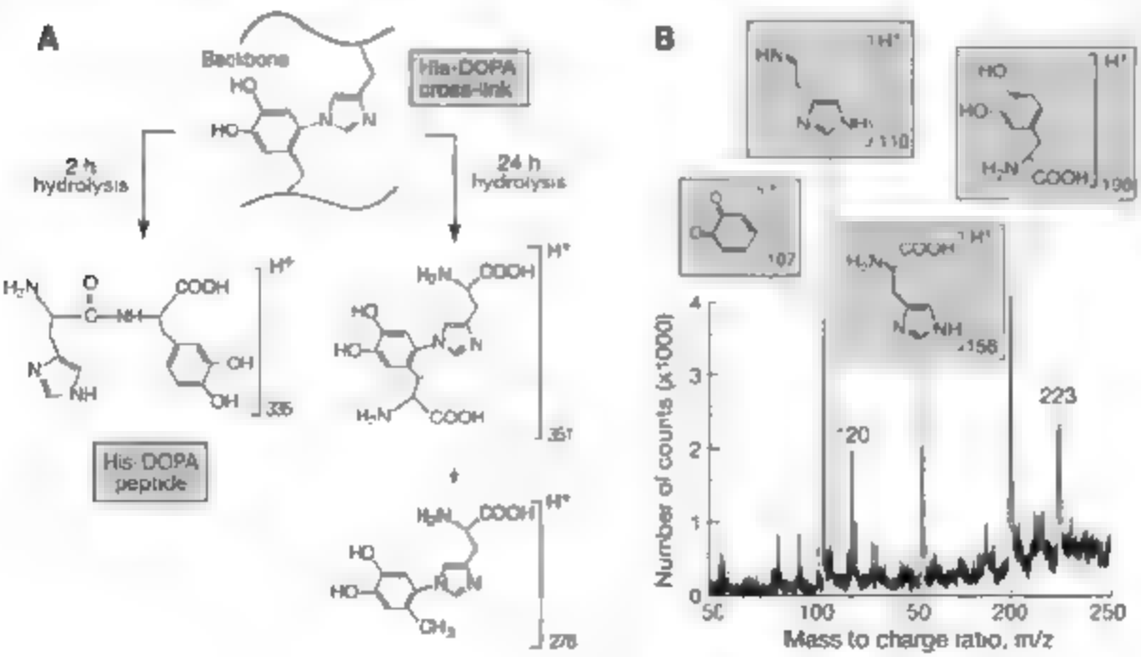


Fig. 3. Dopa and cross-link derivatives. (**A**) His-dopa peptide (*m/z* 335) and His-dopa cross-link (*m/z* 351) captured by phenylboronate chromatography of a hydrolyzed beak. (**B**) Prominent ion fragments in acid-insoluble black pigment, obtained by laser desorption/ionization–time-of-flight mass spectrometry.

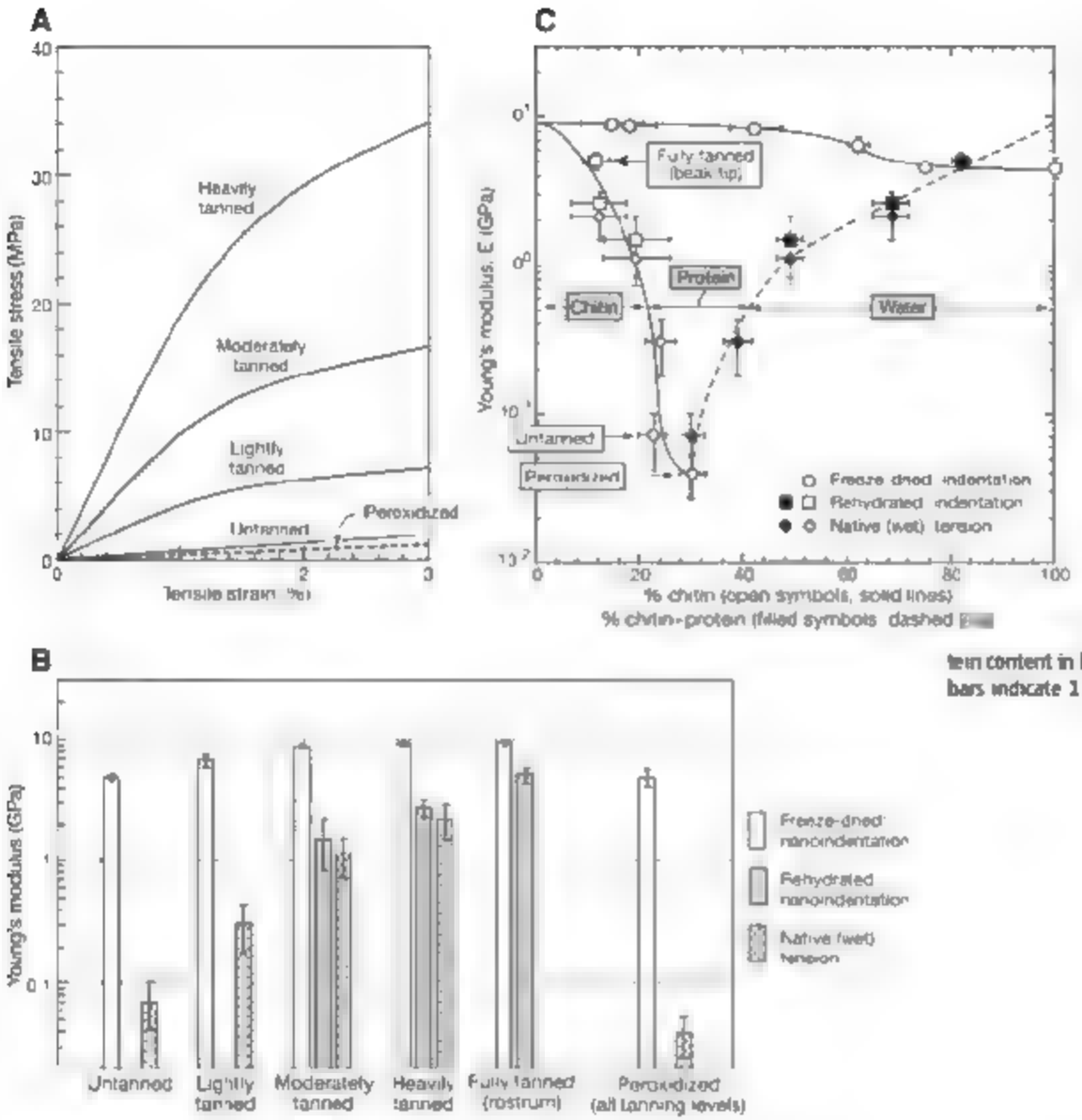


Fig. 4. Mechanical properties of squid beaks. (**A**) Typical tensile stress/strain curves for hydrated beak specimens. Tests could not be performed on the most heavily tanned material at the rostrum because of its size and shape. (**B**) Summary of Young's moduli obtained for dry and hydrated coupons by tensile testing and/or nanoindentation. Nanoindentation tests proved useful for measuring E for all but the two softest states, in untanned and lightly tanned rehydrated samples, the displacements necessary for reliable modulus measurement exceeded the accessible range of the instrument. Tests on peroxidized samples showed no effect of the initial pigmentation level, and thus all data have been combined (far right). Error bars indicate 1 SD. (**C**) Young's modulus plotted against both chitin content and combined chitin/protein content in both dry and hydrated coupons. Error bars indicate 1 SD.

beaks, it is not possible at this time to resolve the magnitude of their respective contributions.

The importance of water in the functional properties of biomolecular materials is widely recognized. Equally important, though subtler, is the suggestion from this study that hydration of wet chitin can be exquisitely tuned from 20 to 70 wt % by association with a cross-linked His-rich protein. Given the available 100-fold range in stiffness, the nature of this association merits further careful study and may offer a new dimension of versatility to bio-inspired dopa-containing polymers (*26*, *27*).

References and Notes

1. M. Benjamin *et al.*, *J. Anat.* **208**, 471 (2006).
2. A. J. Kinloch, *Adhesion and Adhesives: Science and Technology* (Chapman & Hall, London, 1987), pp. 206–209.
3. S. Suresh, *Science* **292**, 2447 (2001).
4. V. Imbeni, J. Kruzic, G. W. Marshall, S. J. Marshall, R. O. Ritchie, *Nat. Mater.* **4**, 229 (2005).
5. D. Raabe, C. Sachs, P. Romano, *Acta Mater.* **53**, 4281 (2005).
6. J. H. Waite, H. C. Lichtenegger, G. D. Stucky, P. Hansma, *Biochemistry* **43**, 7653 (2004).
7. A. M. Sweeney, O. C. Des Marais, Y.-E. A. Ban, S. Johnsen, *J. R. Soc. Interface* **4**, 685 (2007).
8. T. A. Uyeno, W. M. Kier, *J. Morphol.* **264**, 211 (2005).
9. A. Miserez, Y. Li, J. H. Waite, F. W. Zok, *Acta Biomater.* **3**, 139 (2007).
10. R. D. Barnes, *Invertebrate Zoology* (Saunders, Philadelphia, 1966).
11. O. N. Moses, J. D. Harrield, G. D. Stucky, J. H. Waite, *J. Biol. Chem.* **281**, 34826 (2006).
12. T. Nishino, R. Matsui, K. Nakamae, *J. Polym. Sci. Part B: Polym. Phys.* **37**, 1191 (1999).
13. T. L. Hopkins, K. J. Kramer, *Annu. Rev. Entomol.* **37**, 273 [illegible].
14. V. A. Iconomidou, J. H. Willis, S. J. Hamodrakas, *Insect Biochem.* **35**, 533 (2005).
15. J. L. Kerwin *et al.*, *Anal. Biochem.* **268**, 229 (1999).
16. J. Schaefer *et al.*, *Science* **235**, 1200 (1987).
17. A. C. Neville, *Biology of Fibrous Composites: Development Beyond the Cell Membrane* (Cambridge Univ. Press, Cambridge, 1993).
18. J. F. V. Vincent, *Composites Part A* **33**, 1311 (2002).
19. S. O. Andersen, *Annu. Rev. Entomol.* **24**, 29 (1979).
20. C. Sun, J. H. Waite, *J. Biol. Chem.* **280**, 39332 (2005).
21. J. F. V. Vincent, S. Abbott, *J. Insect Physiol.* **33**, 973 (1987).
22. M. Meissner, M. G. Peter, J. F. V. Vincent, *Biomacromolecules* [illegible].
23. N. J. Murray, M. P. Williamson, T. H. Lilley, E. Haslam, *Eur. J. Biochem.* **219**, 923 (1994).
24. P. C. J. Brunet, *Insect Biochem.* **10**, 467 (1980).
25. F. Hook *et al.*, *Anal. Chem.* **73**, 5796 (2001).
26. H. Lee, B. P. Lee, P. B. Messersmith, *Nature* **448**, 338 (2007).
27. B. Liu, L. Burdine, T. Kodadek, *J. Am. Chem. Soc.* **128**, 15228 (2006).

28. Funding was provided by NIH (grants R01 DE014672 and DE015451), a seed grant from the Materials Research Science and Engineering Center Program of NSF under award number DMR05-80034, and NASA University Research Engineering and Technology Institute award number NCC-1-02037. A.M. gratefully acknowledges financial support from the Swiss National Science Foundation through an individual advanced researcher fellowship (grant N° PA002–113176/1). C. Salinas (Centro de Investigaciones Biologicas del Noroeste, La Paz, Mexico), J. C. Weaver, and the crew of the *Corolomo* (Ventura Harbor, CA) helped with squid beak collections. J. C. Weaver and J. Pavlovic provided technical expertise in sample preparation for high-resolution scanning electron microscopy and in operating ESI-MS and MS/MS, respectively.

Supporting Online Material

www.sciencemag.org/cgi/content/full/319/5871/1816/DC1
Materials and Methods
SOM Text
Figs. S1 to S5
References

13 December 2007; accepted 7 February 2008
10.1126/science.1154117

Determining Transition-State Geometries in Liquids Using 2D-IR

James F. Cahoon, Karma R. Sawyer,* Jacob P. Schlegel,* Charles B. Harris†

Many properties of chemical reactions are determined by the transition state connecting reactant and product, yet it is difficult to directly obtain any information about these short-lived structures in liquids. We show that two-dimensional infrared (2D-IR) spectroscopy can provide direct information about transition states by tracking the transformation of vibrational modes as a molecule crossed a transition state. We successfully monitored a simple chemical reaction, the fluxional rearrangement of $Fe(CO)_5$, in which the exchange of axial and equatorial CO ligands causes an exchange of vibrational energy between the normal modes of the molecule. This energy transfer provides direct evidence regarding the time scale, transition state, and mechanism of the reaction.

Transition states determine many properties of chemical reactions, yet there are limited experimental methods to directly probe the details of transition-state geometries and symmetries in liquids. Molecular structures primarily reflect local minima on a potential energy surface, and chemical reactions occur only during brief excursions up and over transition states on these surfaces. The excursions are short-lived, which greatly complicates any measurement of their properties. Most methods with sufficient time resolution rely on photoinitiation of a chemical reaction and provide information about the intermediates formed during the course of the reaction, rather than direct information about transition-state structures (*1*).

Here, we demonstrate that ultrafast two-dimensional infrared (2D-IR) spectroscopy provides evidence for the transition state of a simple thermal chemical reaction, the fluxionality of $Fe(CO)_5$. "Fluxionality" refers to the rearrangement of a molecule between chemically indistinguishable structures. These reactions produce no net change in molecular structure, yet they are important for understanding the basic chemical behavior and reactivity of molecules in solution.

$Fe(CO)_5$, an organometallic complex with five CO ligands arranged in a trigonal bipyramidal geometry, is a textbook example of fluxionality. In the late 1950s, nuclear magnetic resonance (NMR) spectroscopy revealed that this molecule rapidly exchanges its CO ligands between axial and equatorial sites. The ^{13}C NMR spectrum of ^{13}C-labeled $Fe(CO)_5$ exhibits only a single peak at all accessibly measured solution temperatures, which indicates that ^{13}C nuclei shift between axial and equatorial positions faster than NMR spectroscopy is able to distinguish these two chemical environments (*2*–*4*). Careful analysis of IR, Raman, and NMR spectra of $Fe(CO)_5$ and various derivatives suggests that the exchange process possesses a low barrier and occurs on a time scale of picoseconds (*4*–*6*). Nevertheless, the dynamics in solution have not been quantified.

From a general perspective, fluxional processes are simple chemical reactions in which a molecule briefly rearranges to a new symmetry and geometry as it crosses a transition state and then returns to its original geometry as it completes the reaction. We show for $Fe(CO)_5$ that during this process, energy is exchanged between the vibrational modes of the reactive ligands. Quantification of this energy exchange provides direct information about the time scale, the transition state, and, consequently, the mechanism of the reaction.

Much attention has recently been given to 2D-IR spectroscopy because of its ability to monitor thermal reactions and chemical exchange on the femtosecond and picosecond time scales. Conventionally, ultrafast timing of chemical reactions is achieved by photoinitiating the reactions with a short, intense laser pulse, which electronically excites the molecules and typically leaves them with substantial excess energy (*7*). In comparison, 2D-IR spectroscopy only requires vibrational excitation with an ultrafast IR laser pulse and allows the investigation of an entirely different class of thermally activated reactions in liquids at or close to equilibrium. Although similar to 1D-IR pump-probe experiments, 2D-IR spectroscopy separates the contributions to the 1D-IR spectrum into two frequency dimensions, which provides information on the correlations, anharmonicities, and exchange dynamics between all vibrational modes encompassed by the bandwidth of the laser pulses (*7*–*9*). This additional information is necessary to monitor thermal chemical reactions with ultrafast precision.

For example, Fayer and co-workers recently reported direct dynamics measurements of carbon-carbon single-bond rotation (*10*), and in similar experiments several groups have observed the complexation and dissociation of solute-solvent complexes (*11*–*13*). These experiments used 2D-IR spectroscopy to vibrationally tag a specific localized vibrational mode and monitor its shift in frequency as the system underwent chemical dynamics. Here we describe a conceptually different experiment on $Fe(CO)_5$. We imparted vibrational energy into a specific but delocalized vibrational mode and monitored how that energy was distributed between the delocalized modes as the molecule crossed a transition state.

$Fe(CO)_5$ has two IR-active vibrational modes involving the stretch of the CO ligands. For the D_{3h} symmetry of $Fe(CO)_5$, these absorptions correspond to a doubly degenerate e′ band and an a_2'' band, at 1999 and 2022 cm^{-1}, respectively, in *n*-dodecane solution (Fig. 1A). Density functional theory (DFT) vibrational frequency calculations (*14*) yielded CO displacements for the a_2'' and e′ modes (Fig. 1B). The e′ mode involved nearly exclusive vibration of the three equatorial CO groups, whereas the a_2'' mode involved vibration of the axial CO groups.

In principal, fluxionality can cause the coalescence and collapse of the two IR absorptions into a single peak, analogous to the coalescence of line shapes observed in NMR spectra. The frequency of exchange must be comparable to the frequency separation of the absorption bands, which is ~1 ps^{-1} for these absorptions. The Fourier transform IR spectrum at 100°C shows some evidence of coalescence relative to the room temperature spectrum (Fig. 1A), but it is unclear whether the changes in line shape are the result of exchange or homogeneous broadening (see fig. S1). The boiling point (103°C) and thermal instability of $Fe(CO)_5$ prevent the acquisition of spectra at sufficiently high temperatures to conclusively observe IR coales-

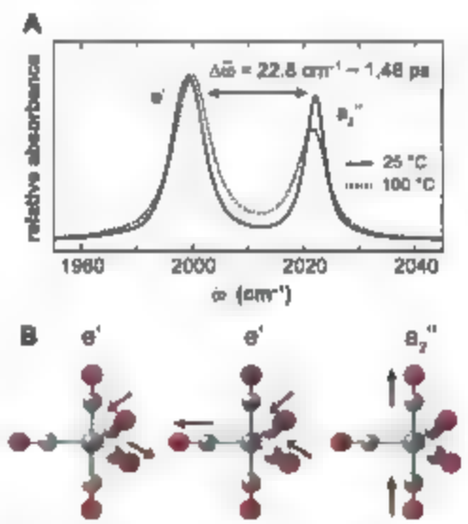


Fig. 1. (**A**) Room-temperature (25°C) Fourier transform IR spectrum (solid line) of $Fe(CO)_5$ in *n*-dodecane and spectrum at 100°C (dotted line). $\bar{\omega}$, the frequency of vibration; $\Delta\bar{\omega}$, frequency separation of the two IR bands. (**B**) Depiction of the approximate eigenvectors for the a_2'' and doubly degenerate e′ vibrational modes of $Fe(CO)_5$.

Department of Chemistry, University of California, Berkeley, CA 94720, USA, and Chemical Sciences Division, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA.

*These authors contributed equally to this work.
†To whom correspondence should be addressed. E-mail: cbharris@berkeley.edu

cence. Nevertheless, several examples of this phenomenon have been reported and attributed to fast exchange (*15*–*17*), although the true physical origin of the effect has been debated in some instances (*18*). The 2D-IR spectra reported below, however, show strong evidence for the time scale of fluxionality and exchange in room-temperature solution and highlight the advantage of 2D-IR as a more general method of observing IR exchange, analogous to the advantages of 2D-NMR over conventional NMR spectroscopy (*19*).

The experiment proceeded as follows. First, a narrowband IR laser pulse selectively excited one of the two vibrational bands in $Fe(CO)_5$, promoting it from the $v = 0$ to $v = 1$ vibrational level. Second, a fixed time elapsed (referred to as the waiting time, T_w) to allow the molecule to exchange axial and equatorial ligands. Third, a broadband IR laser pulse probed the sample, and the resulting absorption spectrum reflected how the vibrational energy imparted by the first pulse had been redistributed between the vibrational modes. The relative polarization between the two laser pulses was set to the magic angle (54.7°) to avoid effects from rotational diffusion. This setup (*20*), based on a pump-probe–style experiment (*21*), gave a time resolution of ~1 ps and eliminated many of the coherence transfer effects that are observed in heterodyne-detected photon-echo 2D-IR experiments (*8*).

2D-IR spectra of $Fe(CO)_5$, normalized to the strongest peak at each T_w, are displayed in Fig. 2 for T_w of 2 ps, 7.5 ps, and 20 ps, and spectra acquired at additional T_w are provided in fig. S3. The vertical axis represents the frequency of the narrowband IR pulse used to selectively excite vibrational modes, and the horizontal axis represents the change in absorbance that resulted from this excitation after T_w. Interpretation of this type of spectrum has been discussed extensively (*7*). Negative signals are shown in blue and appear along the diagonal at (2022, 2022) and (1999, 1999), which correspond to the 0-1 transition frequencies for the a_2'' and e' modes, respectively. Shifted down in frequency along the horizontal axis from the diagonal are positive peaks shown in red, corresponding to the 1-2 transition frequencies. These peaks are shifted off the diagonal by the anharmonicity of the vibrational potential. The 0-1 peaks result both from bleaching of the ground state and stimulated emission from the excited state, whereas the 1-2 peaks result from only $v = 1$ to $v = 2$ transitions (*22*). Unlike 2D-IR experiments on solute-solvent complexes (*12*, *13*), all vibrational modes in these spectra share one common ground-state level, and as a result, the exchange dynamics apparent in both the 0-1 and 1-2 signals reflect only dynamics in the excited vibrational states. The following analysis uses kinetic data from the 0-1 signal because these absorption line shapes are narrower and are clearly defined on the diagonal and off diagonal.

A substantial shift in the peak intensities is observed between $T_w = 2$ and $T_w = 7.5$ ps (Fig. 2B). The most noticeable change is the increase in intensity at the cross peak positions, labeled as triangles in Fig. 2, A to C. Several factors affect the relative intensities of peaks in 2D spectra, and cross peaks typically arise from anharmonic coupling between vibrational modes. To predict the intensity of anharmonic cross peaks between the a_2'' and e' modes, DFT anharmonic frequency calculations (*14*, *23*) were performed and gave an anharmonicity of 0.6 cm^{-1} between the two modes. Because the anharmonic shift is much smaller than the vibrational line widths, the contribution of these peaks to the 2D spectra is present but relatively minor. Vibrational relaxation may also contribute to changes in the peak intensities (*24*), however, vibrational relaxation of the CO stretches occurs on a much longer time scale (~150 ps, fig. S2) than the values of T_w in Fig. 2, and does not affect the interpretation of the

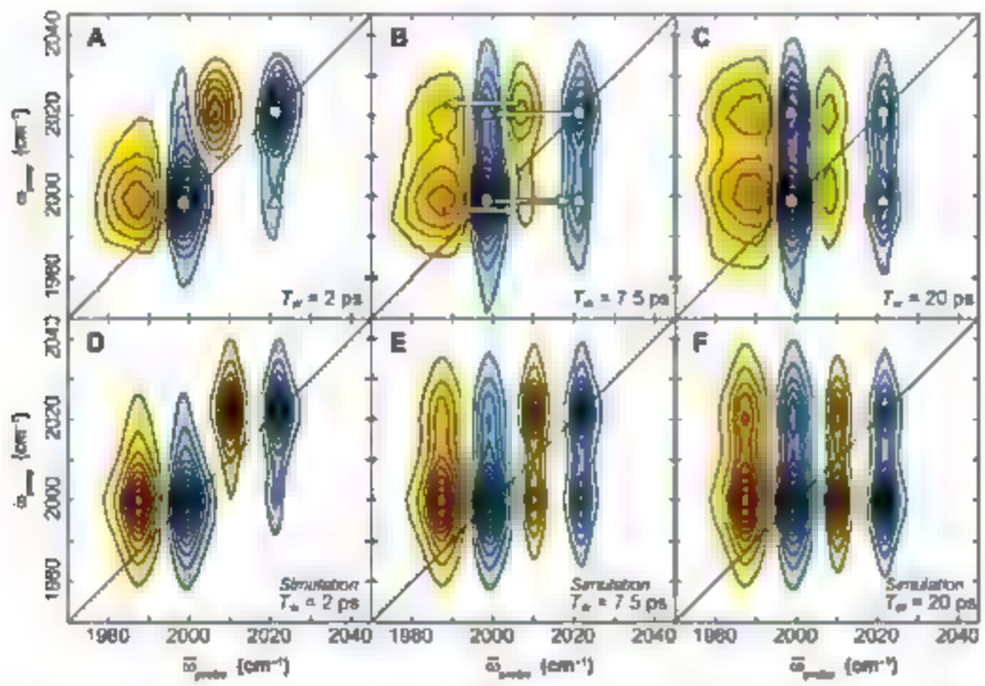


Fig. 2. Contour plots of experimental 2D-IR spectra (top row) for $Fe(CO)_5$ in *n*-dodecane at room temperature (18°C) acquired at T_w's of (**A**) 2 ps, (**B**) 7.5 ps, and (**C**) 20 ps. Simulated 2D-IR data for the same T_w's are shown in (**D**), (**E**), and (**F**), respectively. The data are normalized to the maximum absorption at each T_w, which corresponds to the negative absorption (blue) of the e' vibrational mode (see fig. S2 for the absolute magnitude of this absorption), and contour lines represent a 10% change in absorption intensity. Diagonal peaks are marked by circles and off-diagonal peaks by triangles in (A) to (C). Arrows in (B) represent vibrational energy exchange. Simulations [(D) to (F)] include energy exchange from a pseudo-rotation with a time constant of 8 ps; details of the simulation parameters are provided in the SOM.

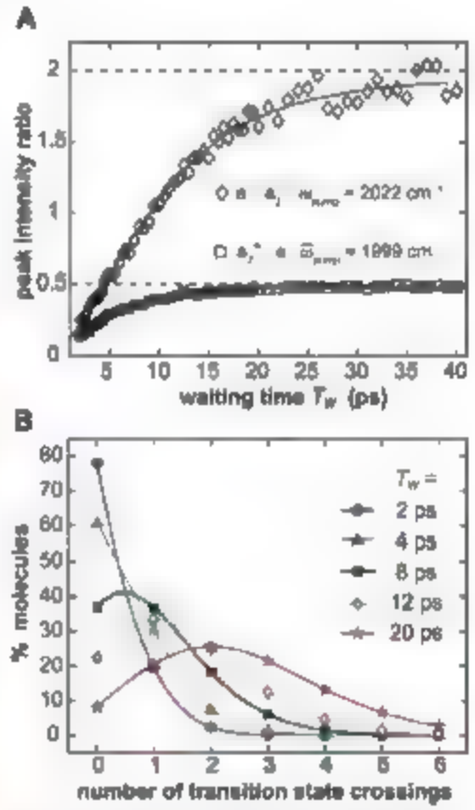


Fig. 3. (**A**) Ratio of 2D-IR off-diagonal–to–diagonal peak intensity at room temperature (18°C) as a function of T_w for selective excitation of the e' band (squares) and a_2'' band (diamonds). Ratios are calculated from Lorentzian fits to 1D horizontal slices of the 2D-IR spectra using the intensities for peaks marked by circles and diamonds in Fig. 2. Solid lines are simulations assuming energy exchange due to the pseudo-rotation; a single time constant of 8 ± 0.6 ps gives the best fit for both curves. (**B**) The distribution of the number of transition-state crossings as a function of T_w for a fluxional process with a kinetic time constant of 8 ps.

spectra, as discussed further in the supporting online material (SOM).

The change in intensity at the diagonal and cross peak positions is attributed to vibrational energy transfer between the CO normal modes as $Fe(CO)_5$ crossed the transition state. Figure 2, D to F, show simulated 2D-IR spectra using DFT-calculated anharmonic parameters (including the cross anharmonicity of 0.6 cm^{-1}) and assuming energy exchange between the vibrational modes. The simulated spectra qualitatively reproduce the experimental data, which indicates that the changes in peak intensity do result from vibrational energy transfer.

The most straightforward way to see the energy exchange in the experimental data is to plot the ratio of off-diagonal (triangles) to diagonal (circles) peak intensities for the two horizontal slices of the 2D-IR spectra at IR pump frequencies ($\bar{\omega}_{pump}$) of 1999 and 2022 cm^{-1} (Fig. 2, A to C). Data were collected for a range of T_w's between 2 and 40 ps, and Fig. 3A displays these ratios as a function of T_w. For excitation of the a_2'' band, the ratio rises to an asymptotic value of 2, and for excitation of the e' band the ratio rises to an asymptote value of 0.5.

In order to quantitatively understand the vibrational energy transfer and asymptotic values shown in Fig. 3A, the details of the molecular rearrangement must be considered. Berry proposed a pseudo-rotation mechanism for this type of trigonal bipyramidal complex. The D_{3h} equilibrium structure rearranges to a transition state with C_{4v} symmetry and then returns to the D_{3h} structure (*5*, *25*). In the process, the two axial ligands are exchanged with two equatorial ligands at the transition state, and the molecule appears to have rotated 90°. We have modeled this mechanism with DFT calculations of the intrinsic reaction coordinate (*14*, *26*), with the results shown in Fig. 4A. The calculations give an electronic energy barrier of 2.13 kcal/mol.

In order to model the vibrational energy transfer observed in the 2D-IR spectra, we first assume that the delocalized normal modes of $Fe(CO)_5$ can be expressed as linear combinations of the individual, localized CO stretch displacements (Eq. 1)

$$|n\rangle = \sum_{j=1}^{5} |CO_j\rangle\langle CO_j|n\rangle \qquad (1)$$

where $|n\rangle$, $n = 1, \ldots 5$, represents the five harmonic normal modes composed of CO stretches and $|CO_j\rangle$, $j = 1 \ldots 5$, represents the displacement vector of the jth CO group. Neglecting changes in this displacement vector as the CO groups change position (i.e., neglecting changes in the local mode force constants and wavefunctions), the initial normal mode $|n_i\rangle$ before rearrangement of the molecule can be expressed as a superposition of the normal modes $|n_f\rangle$ after rearrangement, as in Eq. 2

$$|n_i\rangle = \sum_{n=1}^{5} |n_f\rangle \sum_{j=1}^{5} \langle n_f|CO_j\rangle\langle CO_j|n_i\rangle \qquad (2)$$

where the individual CO groups $|CO_j\rangle$ retain their label j as they rearrange. This analysis was performed, assuming the pseudo-rotation mechanism, by projecting the initial CO normal modes, including the two IR-inactive a_1' modes, onto the C_{4v} transition state and then onto the rearranged D_{3h} structure, using the DFT-calculated vibrations. The results are summarized schematically in Fig. 4B, where the states $|n\rangle$ are labeled by their corresponding symmetries, and coefficients, as calculated from Eq. 2, are provided in front of these labels. The irreducible representations and projections for the vibrations would be different for transition states with different symmetries.

In order to model vibrational energy transfer, we need the probability that the initial vibration $|n_i\rangle$ will transfer energy into a final normal mode $|n_f\rangle$, which is given by Eq. 3

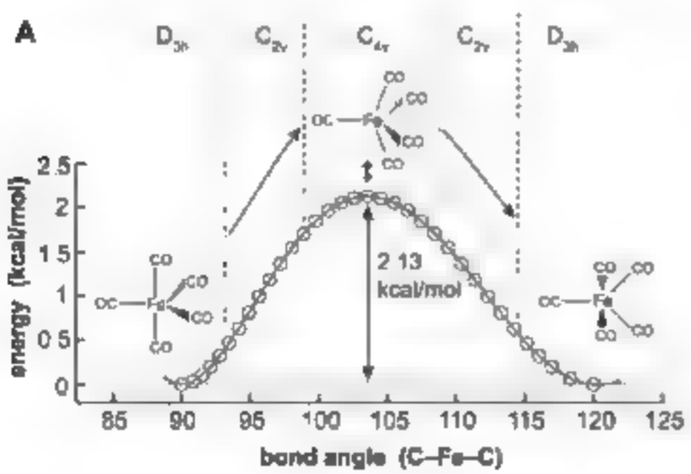
$$P_{n_i \to n_f} = \left| \sum_{j=1}^{5} \langle n_f|CO_j\rangle\langle CO_j|n_i\rangle \right|^2 \qquad (3)$$

These values are given in the far right column of Fig. 4B as percentages. Based on these values, we simulated energy transfer in $Fe(CO)_5$ by using a simple deterministic rate law governed by one rate constant k, the rate of crossing the barrier in the fluxional reaction. The rate of transferring energy from one initial mode into a final vibrational mode is given by Eq. 4

$$k_{n_i \to n_f} = P_{n_i \to n_f} \times k \qquad (4)$$

The kinetics for energy transfer among all the normal modes have been calculated using standard finite difference methods (*27*).

Based on this analysis, 2D-IR data were simulated with energy transfer included as shown in Fig. 2, D to F. In addition, the ratios of off-diagonal to diagonal peak intensity, which directly reflect the relative populations of the e' and a_2'' vibrational modes, were simulated using this model. The best agreement with the experimental data at room temperature (18°C) was achieved



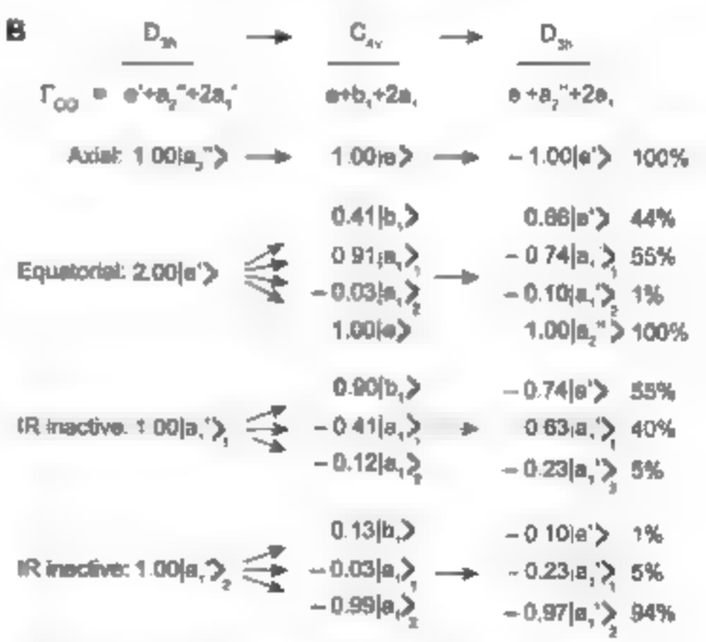


Fig. 4. **(A)** Depiction of the mechanism and energetics of a pseudo-rotation plotted as a function of the bond angle (C–Fe–C between one axial and one equatorial CO ligand that exchange during the pseudo-rotation. Circles are DFT-calculated energies along the intrinsic reaction coordinate, which gives a barrier height of 2.13 kcal/mol. Symmetry labels for the molecule as it passes over the barrier are provided above the structures. **(B)** The splitting of the vibrational modes, denoted by their symmetry labels, as the molecule crosses the transition state. The irreducible representation, Γ_{CO}, for vibrations of the CO groups is given at the top of each column. The individual vibrational modes in the starting structure become a superposition of vibrational modes in the transition state and then again in the final rearranged structure, with the coefficients given for each mode in the superposition. The mixing is limited to those modes that are allowed to mix by symmetry. The final distribution of energy after one crossing over the transition state is shown on the far right in percentages.

with a time constant of 8.0 ± 0.6 ps, or equivalently $k = 1/8$ ps^{-1}, and details of the procedure used to simulate and fit the data are provided in the SOM. The results of the simulation are shown as solid lines through the data in Fig. 3A. The good agreement of the simulations with the experimental data suggests that this model is sufficient to describe the vibrational energy exchange.

Both the simulations and experimental data show that the initial vibrational excitation rapidly reaches an equilibrium between the two IR-active vibrational modes. Because there are two degenerate e′ modes and only one a_2'' mode, this equilibration causes the off-diagonal to diagonal ratios in Fig. 3A to asymptotically approach the values 2 e′ | a_2''. If the molecules crossed the transition state only once, we would expect the asymptotic values instead to reflect the percentages in Fig. 4B. That would be expected for reactions in which the product is thermodynamically stable as compared with the reactant. In this experiment, however, the reactant and product were energetically equivalent and the barrier to the transition state was low, which allowed the molecules to cross the transition state multiple times.

We estimated the distribution of transition-state crossings as a function of T_w, using Gillespie's Monte Carlo algorithm to model the reaction as a stochastic process within the Markov approximation (*27*, *28*). The distribution is displayed in Fig. 3B and shows that within 8 ps, approximately 25% of the molecules crossed the transition state two or more times. Multiple crossings allow the fast exchange of energy between the vibrational modes and cause complete equilibration of the energy within a few crosses of the transition state. The rate of equilibration depends both on the rate of the fluxional reaction and the details of the energy redistribution, which are determined by the transition-state symmetry and, hence, the mechanism of the rearrangement.

Although the pseudo-rotation has been widely accepted as the most likely mechanism for fluxionality in trigonal bipyramidal complexes such as $Fe(CO)_5$, other possibilities have not been ruled out (*29*). The analysis of 2D-IR data described above requires the assumption of a mechanism and a transition-state geometry in order to simulate the energy transfer process. To explore alternative pathways, we simulated data for a different fluxional mechanism in which one axial CO and two equatorial CO groups execute a twist motion, resulting in the exchange of one axial and one equatorial ligand. This mechanism predicts energy exchange dynamics that are different than the pseudo-rotation mechanism, the details of which are provided in the SOM. We found that reasonable agreement with the experimental data at room temperature (18°C) could be achieved only with a much shorter time constant of 3.5 ± 0.4 ps for the twist rearrangement. The difference in time constant between the twist and pseudo-rotation mechanisms demonstrates that the experiment is sensitive to the mechanism, but because of fast multiple crossings, the data at room-temperature alone do not definitively support just one mechanism. Experiments at different temperatures, however, provided the additional data necessary to rule out the twist mechanism.

2D-IR data were collected at two additional temperatures and a kinetic analysis based on a pseudo-rotation was performed at all three temperatures, giving time constants of 8.0 ± 0.6 ps at room temperature (18°C), 6.3 ± 0.8 ps at 50°C, and 4.6 ± 0.4 ps at 90°C. An Arrhenius plot of the data (fig. S5) gives a straight line with an activation energy of 1.6 ± 0.3 kcal/mol. This result is in reasonable agreement with the barrier for the pseudo-rotation mechanism calculated by DFT (2.13 kcal/mol). In contrast, our DFT calculations on the twist mechanism, detailed in the SOM, indicate that the barrier for this type of rearrangement is at least one order of magnitude greater than for the pseudo-rotation mechanism.

The Arrhenius plot not only further confirms the pseudo-rotation mechanism but also demonstrates that the energy transfer results from a barrier-crossing process and not intramolecular vibrational relaxation (IVR) caused by anharmonic coupling between the modes. IVR typically occurs on a time scale of picoseconds, and could potentially compete with fluxionality as the cause of energy exchange between the e′ and a_2'' modes. The rate of IVR depends on the strength of the coupling as well as the existence of liquid phonons (instantaneous normal modes of the solvent) at the frequency of the energy mismatch between the modes (22.8 cm^{-1}), which serve to receive or supply the energy difference (*30*, *31*). At higher temperatures, an increase in the population of low-frequency phonons around 22.8 cm^{-1} may increase the rate of IVR. Because this frequency is low compared with the thermal energy in these experiments (200 to 260 cm^{-1}), the population will increase relatively little with temperature. Between 18° and 90°C, we would expect the IVR rate to increase by ~26% (see SOM for a description of this estimate), yet we observed a change in rate three times greater over this temperature range. IVR is also not supported by the DFT calculations, which predict a weak coupling between these modes (0.6 cm^{-1}) in comparison with other documented examples of IVR (*8*, *32*). Thus, the 2D-IR data strongly support a fluxionality mechanism in which $Fe(CO)_5$ exchanges CO groups and vibrational energy through the C_{4v} transition state of the pseudo-rotation mechanism on a time scale of 8 ps.

When this technique is extended to other systems, the interpretation of the data should be especially straightforward in cases where the system crosses the transition state only once. Even in more complex cases, we anticipate broad applications toward solution-phase transition-state characterization for both thermal and photoinitiated reactions.

References and Notes

1. E. T. J. Nibbering, H. Fidder, E. Pines, *Annu. Rev. Phys. Chem.* **56**, 337 (2005).
2. F. A. Cotton, A. Danti, J. S. Waugh, R. W. Fessenden, *J. Chem. Phys.* **29**, 1427 (1958).
3. R. Bramley, R. S. Nyholm, B. N. Figgis, *Trans. Faraday Soc.* **58**, 1893 (1962).
4. H. W. Spiess, R. Grosescu, U. Haeberle, *Chem. Phys.* **6**, 226 (1974).
5. R. K. Sheline, H. Matrike, *Angew. Chem. Int. Ed. Engl.* **14**, 314 (1975).
6. J. K. Burdett, J. M. Grzybowski, M. Poliakoff, J. J. Turner, *J. Am. Chem. Soc.* **98**, 5728 (1976).
7. M. Khalil, N. Demirdoven, A. Tokmakoff, *J. Phys. Chem. A* **107**, 5258 (2003).
8. M. Khalil, N. Demirdoven, A. Tokmakoff, *J. Chem. Phys.* **121**, 362 (2004).
9. Y. S. Kim, R. M. Hochstrasser, *J. Phys. Chem. B* **111**, 9697 (2007).
10. J. R. Zheng, K. W. Kwak, J. Xie, M. D. Fayer, *Science* **313**, 1951 (2006).
11. S. Woutersen, Y. Mu, G. Stock, P. Hamm, *Chem. Phys.* **266**, 137 (2001).
12. J. R. Zheng *et al.*, *Science* **309**, 1338 (2005).
13. Y. S. Kim, R. M. Hochstrasser, *Proc. Natl. Acad. Sci. U.S.A.* **102**, 11185 (2005).
14. DFT calculations were performed at the BP86 level using the basis sets 6-31+g(d) for C and O and LANL2DZ for Fe with the program Gaussian 03, Revision C.02 (*33*).
15. C. H. Londergan, C. P. Kubiak, *Chem. Eur. J.* **9**, 5962 (2003).
16. F. W. Grevels *et al.*, *J. Am. Chem. Soc.* **120**, 10423 (1998).
17. Y. S. Kim, R. M. Hochstrasser, *J. Phys. Chem. B* **110**, 8531 (2006).
18. H. L. Strauss, *J. Am. Chem. Soc.* **114**, 905 (1992).
19. J. B. Lambert, E. P. Mazzola, *Nuclear Magnetic Resonance Spectroscopy* (Pearson Education, Upper Saddle River, NJ, 2004).
20. Materials and methods are detailed in the supporting material available on *Science* Online.
21. J. Bredenbeck, J. Helbing, P. Hamm, *J. Chem. Phys.* **121**, 5943 (2004).
22. S. Mukamel, *Principles of Nonlinear Optical Spectroscopy* (Oxford Univ. Press, New York, 1995).
23. V. Berone, *J. Chem. Phys.* **122**, 014108 (2005).
24. D. V. Kurochkin, S. R. G. Naraharisetty, I. V. Rubtsov, *Proc. Natl. Acad. Sci. U.S.A.* **104**, 14209 (2007).
25. R. S. Berry, *J. Chem. Phys.* **32**, 933 (1960).
26. C. Gonzalez, H. B. Schlegel, *J. Chem. Phys.* **90**, 2154 (1989).
27. J. I. Steinfeld, J. S. Francisco, W. L. Hase, *Chemical Kinetics and Dynamics*, (Prentice-Hall, Upper Saddle River, NJ, ed. 2, 1989).
28. D. T. Gillespie, *J. Phys. Chem.* **81**, 2340 (1977).
29. I. Ugi, D. Marquard, H. Klusacek, P. Gillespie, F. Ramirez, *Acc. Chem. Res.* **4**, 288 (1971).
30. H. J. Bakker, *J. Chem. Phys.* **98**, 8496 (1993).
31. A. Tokmakoff, B. Sauter, M. D. Fayer, *J. Chem. Phys.* **100**, 9035 (1994).
32. J. D. Beckerle, M. P. Casassa, R. R. Cavanagh, E. J. Heilweil, J. C. Stephenson, *Chem. Phys.* **180**, 487 (1992).
33. M. J. Frisch *et al.*, Gaussian 03, Revision C.02 (Gaussian, Wallingford, CT, 2004).

34. We thank P. Hamm and J. Helbing for recommendations on the experimental apparatus. This work was supported by the NSF's Division of Physical Chemistry. We also acknowledge some specialized equipment supported by the U.S. Department of Energy, Office of Basic Energy Sciences, Chemical Sciences Division, under contract DE-AC02-05CH11231 and contractor supported research. J.F.C. acknowledges an NSF graduate research fellowship.

Supporting Online Material

www.sciencemag.org/cgi/content/full/319/5871/1820/DC1

Materials and Methods
SOM Text
Figs. S1 to S5
References

11 December 2007; accepted 25 February 2008
10.1126/science.1154041

Surface Trapping of Atoms and Molecules with Dipole Rings

Hugo Dil,[1,2] Jorge Lobo-Checa,[1,2]* Robert Laskowski,[3] Peter Blaha,[3] Simon Berner,[1] Jürg Osterwalder,[1] Thomas Greber[1]†

The trapping of single molecules on surfaces without the formation of strong covalent bonds is a prerequisite for molecular recognition and the exploitation of molecular function. On nanopatterned surfaces, molecules may be selectively trapped and addressed. In a boron nitride nanomesh formed on Rh(111), the pattern consisted of holes 2 nanometers in diameter on a hexagonal superlattice, separated by about 3 nanometers. The trapping was further investigated with density functional theory and the photoemission of adsorbed xenon, where the holes were identified as regions of low work function. The analysis showed that the trapping potential was localized at the rims of the holes.

Although molecules can be trapped at surfaces by the formation of strong chemisorption bonds, for many applications as well for fundamental studies it is useful to trap molecules at specific parts of a surface through weak interactions that minimally perturb the bonding within the molecule. Surface dipoles, which contribute strongly to the bonding of polarizable entities, should be able to laterally immobilize molecules if they also exhibit in-plane components.

Electrostatic trapping caused by polarization-induced bonding offers a promising way to immobilize a molecule with minimal coupling to its support and neighbors. For atomic steps and kinks on surfaces, it is known that the lateral variation of the surface potential leads to strong electric fields parallel to the surface that may trap molecules (*1*, *2*). There are also recent reports on the immobilization and trapping of molecules with a chlorine center on defects of potassium bromide supports (*3*). Often, however, it is desirable to build structures that do not rely on random defects. Dislocation networks are periodic arrays of defects (*4*) on which molecules may be trapped. In this regard, hexagonal boron nitride (h-BN) nanomesh, a robust superstructure with a 3.2-nm lattice constant (*5*), has emerged as a promising basis for building nanostructures because it creates a ring of in-plane dipoles. For example, if naphthalocyanine, a molecule with a diameter of about 2 nm, is evaporated on the nanomesh, isolated molecular entities separated by 3.2 nm are found immobilized at room temperature (*6*). Nevertheless, not much is known about the nature of the trapping. Here, we propose a polarization-induced trapping mechanism on the basis of the observation of Xe conglomerates on the nanomesh.

It is now established that h-BN nanomesh on Rh(111) or Ru(0001) is a single sheet of h-BN, where 13 × 13 BN units form a coincidence lattice with 12 × 12 substrate unit cells (*6*–*8*). Two distinct BN regions were found: A closely bound region assigned to the "holes" or "pores" in the nanomesh and a loosely bound region assigned to the "wires" (*5*, *8*). Site-selective bonding of the BN nitrogen units to the substrate atoms creates a corrugation with an amplitude of about 0.05 nm and strongly affects the electronic structure (*8*). Theory confirmed the experimentally observed electronic structure (*5*–*7*) and showed the sp^2-derived σ-band density of states of the holes and the wires to be quasi-rigidly shifted with respect to each other by about 1 eV (*8*).

This finding suggests that a substantial portion of the h-BN σ-band shift is related to electrostatics; that is, to different work functions of the wires and the holes. The accompanying local vacuum-level misalignment imposes electric fields and leads to lateral polarization within the h-BN sheet. Heuristically, this structure may be described with dipole rings, which in turn offer a natural explanation for an enhanced immobilization of molecules. Permanent or induced dipoles in molecules are expected to have an affinity for the largest electrostatic potential gradients. The concept of in-plane dipoles at surfaces is reminiscent of the Smoluchowski effect on metallic surfaces, where delocalized valence electrons on steps or kinks cause the polarization (*9*). In the present case, the dipoles are formed by the contact of two regions with different work functions.

At room temperature, phthalocyanines (Pc's) diffuse on most flat surfaces and form islands (*10*, *11*). Instead, on the h-BN nanomesh, the molecules are isolated and trapped in the holes. Room-temperature scanning tunneling microscopy (STM) images of h-BN nanomesh with Cu-Pc molecules are shown in Fig. 1 (*12*). The van der Waals diameter of these molecules is 1.5 nm and thus slightly smaller than the 2-nm holes of the nanomesh. The 3.2-nm periodicity of the nanomesh is visible, where at these tunneling conditions the wires map dark (*13*). The brightest objects are identified as single molecules trapped in the holes, although at room temperature they are not completely immobilized, which is reflected in the smearing of the molecular images. An analysis of the center of gravity of 11 molecules in Fig. 1A shows them to be randomly shifted by 0.16 ± 0.1 nm off the center of the corresponding host hole as indicated in the insets.

In order to explore the energy landscape of the nanomesh, we performed photoemission experiments on adsorbed Xe. The bonding of Xe to surfaces resembles in many aspects that of molecules such as Cu-Pc. Thus the Xe electron binding energies and the Xe desorption temperatures provide direct insight into the interplay between the surface electronic dipoles and the bonding caused by polarization. The electrostatic landscape in front of a surface is reflected in the Xe core-level binding energy that aligns to the local vacuum level at the atom core position (*14*–*17*), whereas the Xe desorption temperatures indicate the local bonding strength (*18*).

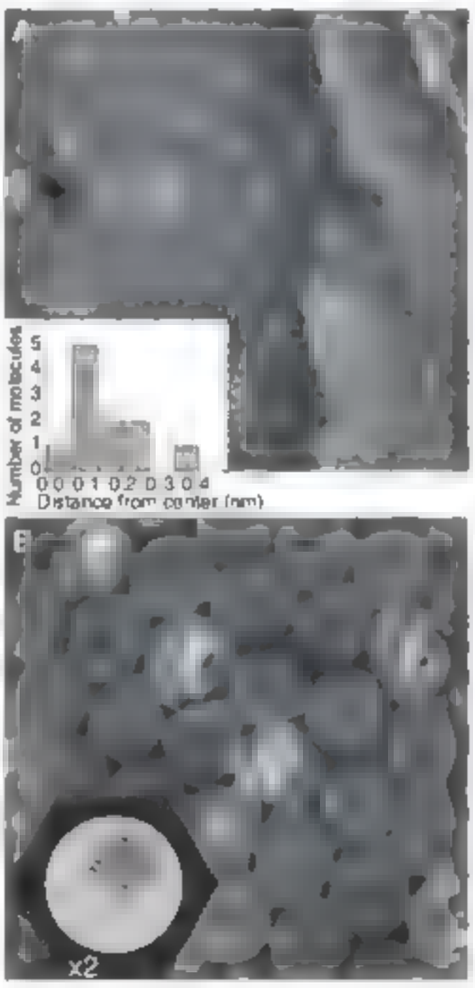


Fig. 1. Room-temperature STM images of a low coverage of Cu-Pc ($C_{32}H_{16}CuN_8$) trapped in the holes of the nanomesh. At the given tunneling conditions, the mesh wires map dark and the holes map bright. Cu-Pc molecules are imaged as bright objects. Current = 30 pA, bias voltage = 1.2 V; images were obtained by horizontal scanning at a rate of 1.8 Hz. (**A**) Overview image (50 by 50 nm). Eleven of 300 nanomesh holes were occupied by molecules. Two monoatomic steps of the Rh substrate and a defect are resolved. The inset shows the distribution of the distances of the molecular center of gravity from the hole center. (**B**) High-resolution image (18 by 18 nm). The inset shows a magnified model of the trapped Cu-Pc molecule.

[1]Physik-Institut, Universität Zürich, Winterthurerstrasse 190, CH-8057 Zürich, Switzerland. [2]Swiss Light Source, Paul Scherrer Institute, CH-5232 Villigen, Switzerland. [3]Institute of Materials Chemistry, Vienna University of Technology, Getreidemarkt 9/165-TC, A-1060 Vienna, Austria.

*Present address: Department of Physics, University of Basel, Klingelbergstrasse 82, CH-4056 Basel, Switzerland. †To whom correspondence should be addressed. E-mail: greber@physik.uzh.ch

Figure 2A displays a series of normal emission spectra obtained during the controlled temperature rise that leads to desorption of a multilayer of Xe on the nanomesh. Four different regimes can be discriminated: the multilayer regime below 63 K, a region between 63 and 72 K [coexistence (C) phase], where two Xe 5p doublets are present; followed by a regime up to 81 K with only one doublet [ring (R) phase]; and above this temperature no Xe features are visible. Energy distribution curves obtained along the dashed lines in Fig. 2A are representative for each of these regions and are plotted in Fig. 2B. The spectrum at 82 K shows the bare nanomesh with the σ_α band at 4.5 eV, which reflects the wires (W), and the σ_β band (at 5.6 eV) associated with the holes (H) (*4*, *7*). The spectrum at 75 K represents the R phase. It shows one Xe $5p_{3/2}$ 5p $_{1/2}$ doublet and the nanomesh σ_α (wire) band. Taking into account that the short mean free path of photoelectrons strongly damps the signal from the substrate if it is covered with a single layer of Xe, we assign this doublet (H) to Xe in the holes of the nanomesh (Fig. 2C). At 66 K, in the C phase, two 5p doublets labeled H and W were observed. The doublet W corresponds to Xe on the wires (Fig. 2D), because its disappearance coincides with the appearance of the σ_α (wire) band.

From the van der Waals radius of Xe of 0.22 nm, we estimate the monolayer coverage to be about 54 ± 1 Xe atoms per nanomesh unit cell. If we assume proportionality of Xe coverage and photoemission intensity, this value corresponds to 25 Xe atoms on the wires and 29 Xe atoms in the holes of the nanomesh unit cell. Some Xe in the holes desorbs at higher temperatures and is bound more strongly.

In the *C* phase, the difference in photoemission binding energy between the Xe adsorbed on H and W sites on the substrate is 310 ± 5 meV. We assign these states to regions of different local work functions, where that on the wires is greater than that in the holes. This difference in energy is in good agreement with the theoretical prediction of 0.5 eV (Fig. 3A) (*19*). Extrapolating the Xe $5p_{3/2}$ binding energy to zero coverage, and using the work function as determined from the secondary electron cutoff of 4.15 eV, we obtain a binding

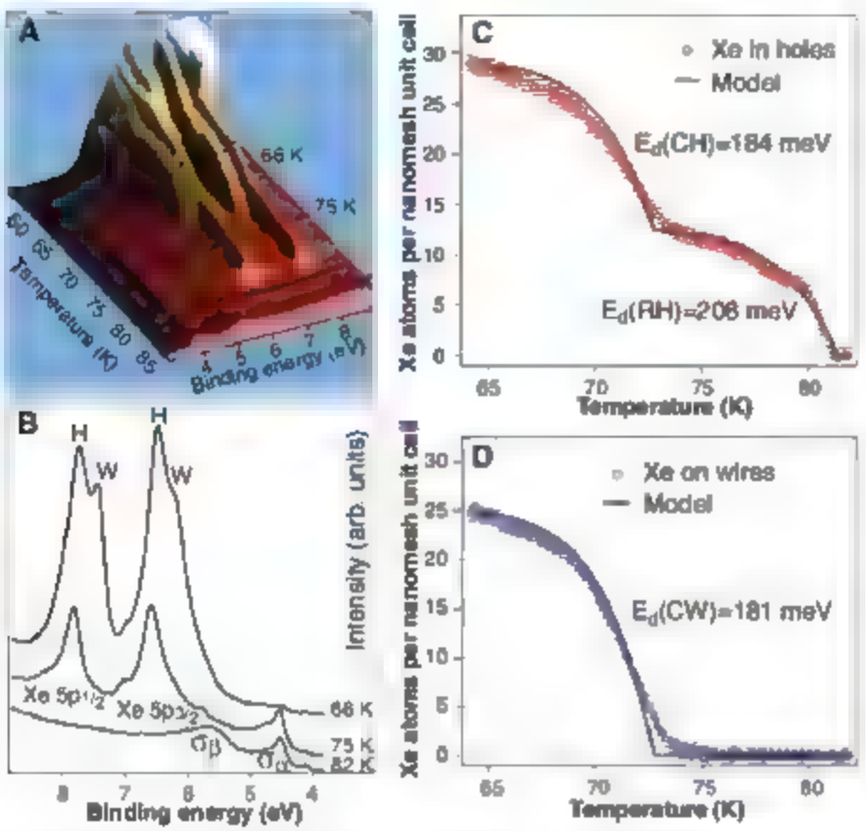


Fig. 2. Temperature dependence of normal emission valence band photoemission spectra from Xe/h-BN/Rh(111). (**A**) Three-dimensionally rendered data set of the desorption of Xe. The dashed lines indicate where the spectra in (B) are derived from. (**B**) Energy distribution curves extracted for three different temperatures (66, 75, and 82 K) as indicated by the dashed lines in (A). (**C**) Spectral weight of the Xe in the holes as a function of temperature. E_d(CH) and E_d(RH) are the two distinct desorption energies from the holes. (**D**) Spectral weight of the Xe on the wires as a function of temperature. E_d(CW) is the desorption energy of Xe from the wires. The solid lines in (C) and (D) are fits obtained from a zero-order desorption model.

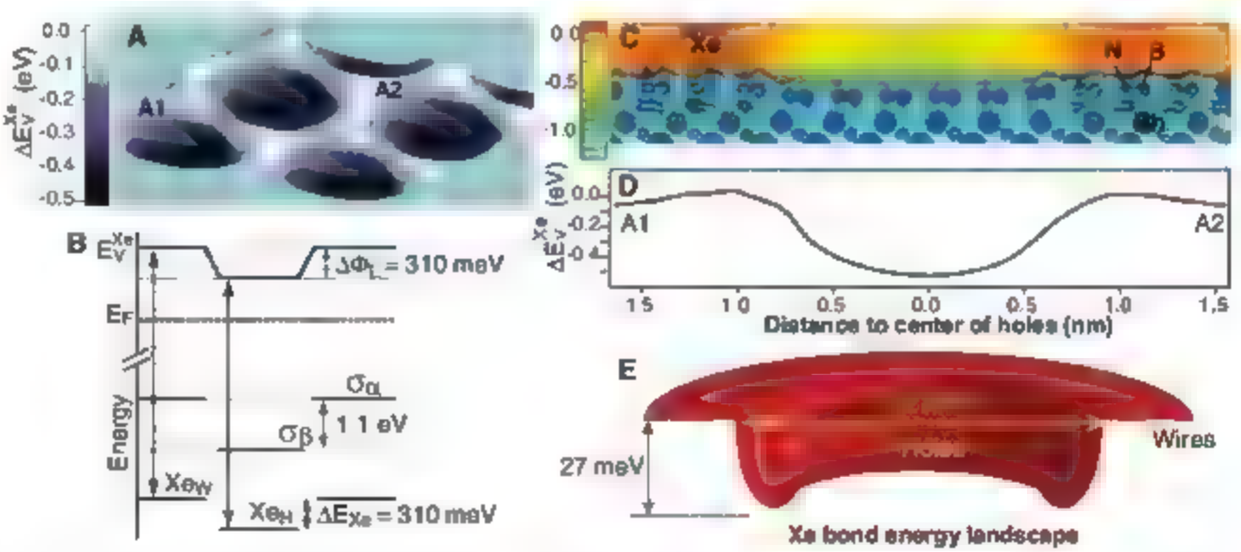


Fig. 3. Energy landscapes of the nanomesh. (**A**) Calculated electrostatic potential 0.38 nm above the nitrogen atoms. The electrostatic potential above the wires is taken as a reference. The potential in the holes attracts negative charge and is 0.5 eV lower. The potential varies because of locally different charge transfer from the h-BN to the substrate and reflects the dipole rings. (**B**) Schematic representation of the observed photoemission fine energies from Xe and the h-BN σ bands on the wires (W) σ_α and in the holes (H) σ_β, respectively. The binding energy difference between Xe_W and Xe_H, ΔE_{Xe}, reflects a difference in the electrostatic potential $\Delta\Phi_L$ at different Xe core positions E_V^{Xe}. (**C**) Vertical cut (5.58 by 1.1 nm) across the unit cell containing points A1 and A2 [see (A)]. The color code and the contours reflect the peculiar electrostatic potential energy landscape. The depicted van der Waals contour of the Xe atom is placed at a site on wires where $E_V^{Xe} = = 0$. (**D**) Electrostatic potential along a line connecting A1 and A2, 0.38 nm above the nitrogen atom cores. (**E**) Bond energy landscape for Xe as derived from the thermal desorption data in Fig. 2, C and D. We propose that the strongest bonding sites are related with the dipole rings at the rims of the holes.

energy relative to the vacuum level of 12.0 eV. This value is between those observed for the first and second layer of Xe on metals (*20*) and indicates that the h-BN sheet has a nonmetallic dielectric response similar to that of one layer of Xe (*21*, *22*).

The photoemission binding energies are summarized in Fig. 3B. It rationalizes the peculiar vacuum energy landscape of the nanomesh, where the holes are local low-work-function patches with a diameter of 2 nm. E_V^{Xe} is the electrostatic potential at the Xe atom cores and varies within the nanomesh unit cell. Figure 3A shows the calculated electrostatic potential at the position of the Xe cores, 0.38 nm above the nitrogen atom cores. The potential has a corrugation amplitude of 0.5 eV, which is reflected in the Xe 5p photoemission final state energies. In Fig. 3, C and D, the potential is shown along the two nonequivalent points A1 and A2 in the nanomesh unit cell. The lower potential in the holes is a consequence of stronger hybridization and concomitant charge transfer from the h-BN mesh to the substrate. It causes lateral electric fields, where electrostatic lensing and an influence on the bonding can be expected.

In order to explore the bonding strength, the data in Fig. 2 were analyzed with kinetic desorption models. The Xe bond energy difference is not expected to be as large as the 310 meV for electrons, because Xe atoms are closed-shell and desorb as neutral species. In the following discussion, the temperature dependence of the Xe desorption from the nanomesh is used to confirm the site-dependent bonding and to corroborate the dipole ring hypothesis.

The peak intensities obtained from the fitting of two Gaussians to the peaks labeled H and W in the Xe $5p_{3/2}$ spectrum are shown in Fig. 2, C and D. All Xe on the wires desorbs at lower temperatures, whereas some Xe in the holes remains longer on the surface. If we take the bond energy to be proportional to the maximum desorption temperature, we find that the Xe atoms on the wires are 15% and those in the holes up to 27% more strongly bound than Xe on Xe (multilayer desorption).

Thermal desorption from the holes occurs both in the C phase, where H-Xe and W-Xe coexist (see spectrum at 66 K in Fig. 2B), and in the R phase, without W-Xe. In the C phase, the ratio between H-Xe and W-Xe is almost constant. An Arrhenius analysis shows that the bond energy difference of the desorbing W-Xe and the H-Xe is less than 5 meV. Xe in the R phase is located in the holes and has a higher bond energy. At 73 K, the R phase comprises about 12 Xe atoms per hole, which corresponds to a compact island with a diameter of 1.6 nm, or to a ring with an outer diameter of 2.1 nm. In order to quantitatively access the Xe bond energies, the data were compared to zero-order kinetics. The desorption rates $-dN/dt$ of a given component are

$$\frac{dN}{dt} = \nu N^0 e^{-E_d / kT} \qquad (1)$$

where ν is an attempt frequency in the order of Xe vibration frequencies ($2 \times 10^{11}\ s^{-1}$) (*23*) and E_d is the desorption energy. The thermal energy kT is related via the heating rate $\beta = dT/dt$ to the desorption rate. With the corresponding initial coverages, integration of Eq. 1 then leads to $N(T)$.

This model with E_d as a single free parameter fits the observed desorption for all three components quite well. In the C phase, we obtain E_d(CW) = 181 meV and E_d(CH) = 184 meV for the holes and wires, respectively, and in the R phase E_d(RH) = 208 meV. The relative accuracy of these values is better than 1%. The results for the C phase confirm a small energy difference between Xe in the holes and on the wires. The average value is near the bond energy of Xe in the second layer on Ru(0001) (169 meV) (*15*) and confirms that one layer of Xe on Ru(0001) produces a similar bond to Xe as one monolayer of h-BN on Rh(111). On the other hand, the R phase exhibits substantial extra bonding of the Xe.

The question of where inside the hole the atoms of the R phase are located is not directly answered by the desorption kinetics. We see no other physical cause for the extra bonding than the dipole rings, which are located at the rims of the holes. This model is in line with the similar Xe bonding in the C phase for Xe on the wires and some of the Xe in the holes. Also, the size of a Xe_{12} ring exactly fits to the rim of a 2-nm hole, and the tendency of Cu-Pc to sit in the hole, but off center, supports this assignment. Furthermore, the extra bond energy of 24 meV exceeds that of a single Xe-Xe bond (20 meV), which in turn gives an argument for a stabilization of an open structure with low-coordinated atoms, despite the higher coordination within a compact island. Figure 3E shows a schematic drawing of the Xe bond energy landscape on the nanomesh, with extra bonding at the rim of the nanomesh holes, where the electrostatic potential gradient is largest. The polarization-induced increase in bond energy is expected to depend on the size and polarizability of the molecule. The in-plane polarizability of Cu-Pc is more than a factor of 30 greater than that of Xe (*24*), and thus provides an explanation of the room-temperature trapping of large polarizable molecules.

The strong bonding sites are proposed to lie on the rim of the nanomesh holes and are reminiscent of the quantum corrals built atom by atom by Eigler *et al.* (*25*, *26*). Our findings suggest that such seemingly artificial structures may also form or be preformed at once and in huge quantities in a self-assembly process on a suitable template such as the nanomesh. The vacuum energy landscape of the nanomesh, which is imposed by dipole rings, gives a physical basis for such a functionality. This concept should also explain the trapping of larger molecules in periodic arrays on a single-layer dielectric at room temperature. We expect that dipole rings will be found in other nanostructures and that they are a quite general scheme to obtain the immobilization of molecules on surfaces.

References and Notes

1 K. Heinann, B. Gumhalter, K. Wandelt, *Surf. Sci.* **251/252**, 289 (1990).
2 M. Morgenstern, T. Michely, G. Comsa, *Phys. Rev. Lett.* **77**, 703 (1996).
3 L. Nony *et al.*, *Nano Lett.* **4**, 2185 (2004).
4 H. Brune, M. Giovannini, K. Bromann, K. Kern, *Nature* **394**, 451 (1998).
5 M. Corso *et al.*, *Science* **303**, 217 (2004).
6 S. Berner *et al.*, *Angew. Chem. Int. Ed.* **46**, 5115 (2007).
7 A. Gorstchko *et al.*, *Langmuir* **23**, 2928 (2007).
8 R. Laskowski, P. Blaha, Th. Gallauner, K. Schwarz, *Phys. Rev. Lett.* **98**, 106802 (2007).
9 R. Smoluchowski, *Phys. Rev.* **60**, 661 (1941).
10 K. Jung Lu, W. Hipps, X. D. Wang, U. Mazur, *J. Am. Chem. Soc.* **118**, 7197 (1996).
11 J. Åhlund *et al.*, *Surf. Sci.* **601**, 3661 (2007).
12 Details of the experimental studies and theoretical calculations are available as supporting material on *Science* Online.
13 R. Laskowski, P. Blaha, *J. Phys. Cond. Mat.* **20**, 064207 (2008).
14 K. Wandelt, *J. Vac. Sci. Technol. A* **2**, 802 (1984).
15 J. Kuppers, K. Wandelt, G. Ertl, *Phys. Rev. Lett.* **43**, 928 (1979).
16 N. D. Lang, A. R. Williams, *Phys. Rev. B* **25**, 2940 (1982).
17 Photoemission final state effects, that is, a site dependence of the photo-hole screening, also have to be considered, although within the same layer they can be treated as a constant offset between the local vacuum-level binding energy and the gas-phase Xe 5p final state.
18 W. Widdra *et al.*, *Phys. Rev. B* **57**, 4111 (1998).
19 The theoretical work functions of approximate structures based on the commensurate 1 × 1 model, with BN at face-centered cubic (fcc, top) and hexagonal close-packed (hcp sites above Rh are 3.7 and 4.0 eV, respectively. The full nanomesh simulations, however, give an average work function of 3.85 eV with a difference of 0.5 eV between the hole and wire regions. The theoretical predictions are in good agreement with experiment, although theory does not consider different final-state screening.
20 G. Kaindl, T.-C. Chiang, D. E. Eastman, F. J. Himpsel, *Phys. Rev. Lett.* **45**, 1808 (1980).
21 Considering that the final-state screening decreases with distance from the metal support, we expect a small increase in Xe 5p binding energy (about 50 meV) in climbing the 0.05 nm from the holes to the wires. This effect would lead to a shift to higher Xe-W binding energy, opposite to what is observed here. The photoemission data also show that the d_{z^2}–d_{π} splitting of 1.1 eV is not only related to a work function difference, as is expected for a weakly physisorbed system (*22*), but points to differences in the electronic structure of the h-BN wire and hole regions.
22 A. Nagashima, N. Tejima, Y. Gamou, T. Kawai, C. Oshima, *Phys. Rev. Lett.* **75**, 3918 (1995).
23 G. Kerner, O. Stein, Y. Lilach, M. Asscher, *Phys. Rev. B* **71**, 205414 (2005).
24 R. Ramprasad, N. Shi, *Appl. Phys. Lett.* **88**, 222903 (2006).
25 D. M. Eigler, E. K. Schweizer, *Nature* **344**, 524 (1990).
26 E. J. Heller, M. F. Crommie, C. P. Lutz, D. M. Eigler, *Nature* **369**, 464 (1994).

27 The photoemission measurements were performed at the Swiss Light Source, Paul Scherrer Institut, Villigen, Switzerland. We thank the staff at the Surface and Interface Spectroscopy beamline at the Swiss Light Source for assistance. The project was supported by the European Commission under contract NMP4-CT 2004-013817 "NanoMesh" and the Swiss National Science Foundation, contract 200020-116096. Technical support by M. Klöckner, F. Dubi, and C. Hess is gratefully acknowledged.

Supporting Online Material

www.sciencemag.org/cgi/content/full/319/5871/1824/DC1
Materials and Methods
References

14 December 2007; accepted 14 February 2008
10.1126/science.1154179

Nutritional Control of Reproductive Status in Honeybees via DNA Methylation

R. Kucharski,* J. Maleszka,* S. Foret, R. Maleszka†

Fertile queens and sterile workers are alternative forms of the adult female honeybee that develop from genetically identical larvae following differential feeding with royal jelly. We show that silencing the expression of DNA methyltransferase Dnmt3, a key driver of epigenetic global reprogramming, in newly hatched larvae led to a royal jelly–like effect on the larval developmental trajectory; the majority of Dnmt3 small interfering RNA–treated individuals emerged as queens with fully developed ovaries. Our results suggest that DNA methylation in *Apis* is used for storing epigenetic information, that the use of that information can be differentially altered by nutritional input, and that the flexibility of epigenetic modifications underpins, profound shifts in developmental fates, with massive implications for reproductive and behavioral status.

Many organisms respond to environmental conditions by displaying phenotypic plasticity, that is, producing different phenotypes from the same DNA genome (*1*, *2*). In social insects, the production of contrasting adult morphologies as well as different reproductive and behavioral systems is critical to their social organization and division of labor (*3*–*6*). These differences are likely to arise at the most basic level from differential somatic imprinting during development on the same genome. Honeybees (*Apis mellifera*) differentially feed genetically identical female larvae to create mainly workers and, when required, a few queens (*7*–*11*). Young nurse bees in the hive produce and feed a largely biochemically uncharacterized substance named royal jelly to larvae destined to become queens, whereas the other larvae are fed with less-sophisticated food (*8*–*11*). Despite their identical clonal nature at the DNA level, workers and queens differ markedly in morphological and physiological features and have contrasting reproductive capabilities, strikingly diverse life spans, and very different behavioral repertoires. It is not understood, however, how differential nutrition is linked to gene expression and how alterations in diet alter pathways that modify the developmental trajectory of an organism.

Studies in mammals suggest that environmental stimuli such as diet can alter the epigenetic state of the genome and affect gene expression by modifying DNA methylation and histone acetylation patterns (*12*, *13*). In addition, strong epidemiological data reveal that cardiovascular and diabetes mortality in children can be influenced by the nutritional status of their parents and grandparents (*14*). In experimental mammalian contexts such as the *agouti* mouse, a number of contrasting phenotypes, such as yellow and obese or brown and slim, can be controlled by varying the mother's diet before, during, and after pregnancy (*15*). *Agouti* gene expression can be silenced by DNA methylation, and its quantities are variable in genetically identical individuals because of epigenetic modifications established during early development (*16*). Differential maternal behavior in the rat also alters the methylation status of the promoter of the glucocorticoid receptor of her pups (*17*).

Epigenetic regulation thus facilitates the integration of intrinsic signals and environmental signals by using highly conserved enzymatic machinery

Molecular Genetics and Evolution, ARC Centre for the Molecular Genetics of Development, Research School of Biological Sciences, Australian National University, Canberra ACT 0200, Australia

*These authors contributed equally to this work.
†To whom correspondence should be addressed. E-mail: [redacted]

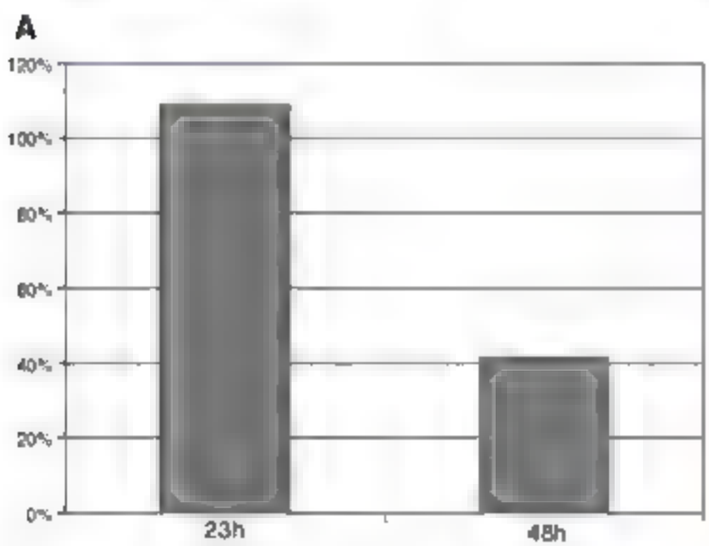


Fig. 1. (**A**) Injections of *Dnmt3* siRNA induce a significant down-regulation of larval *Dnmt3* mRNA levels at stage L3. *Dnmt3* expression was measured in pooled larvae with use of quantitative PCR as described previously and shown relative to a reference gene encoding calmodulin as detailed previously [see (*21*) for details]. (**B**) In situ hybridization showing the expression of *Dnmt3* during L1 to L3 developmental stages. Red arrows indicate the position of the CA. Only the larval heads are shown. The in situ hybridizations were conducted on larvae from a different silencing experiment than the experiment for quantitative PCR.

that includes DNA cytosine-5-methyltransferases (Dnmts), histone deacetylases (HDAs), and methyl-binding proteins (MBPs) (*18*). Although Dnmts appear to be widely conserved in evolution, evidence for a fully functional DNA methylation system in insects has not been forthcoming. Recently, we reported that the honeybee has a full complement of all three functional Dnmts with in vivo properties similar to those of the CpG methylation system in vertebrates (*19*). *Apis* has two orthologs of Dnmt1, one ortholog of Dnmt2, and one of Dnmt3 (*19, 20*). However, dipterans, such as *Drosophila* and mosquitoes, lack some members of this family and possess only a Dnmt2 ortholog. The *Apis* genome also encodes conserved MBPs including components of the nucleosome remodeling and HDA complex. Because all three Dnmt enzymes are shared by humans and honeybees but not by other commonly used model invertebrates, such findings establish the honeybee as a model to not only study the function of DNA methylation in invertebrates, but also for examining any fundamental overlaps that may help in understanding the nutritional basis of epigenetic reprogramming in humans. Although the global methylation landscape of the *Apis* genome remains ill-defined, several methylated genes have already been identified, revealing that CpG methylation occurs preferentially within the coding exons but not at the 5′ and 3′ regions of the transcription units (*19*).

Accordingly, we examined how this highly conserved molecular machinery in *Apis* could be used to elucidate the contrasting anatomical, physiological, and behavioral characteristics of honeybee castes that are brought about by differences in their early life environment.

We used RNA interference (RNAi) technology to silence the expression of Dnmt3 in newly hatched L1 larvae and in embryos because this de novo methylase has been implicated in establishing DNA methylation patterns during development (*18, 20*). Injections of Dnmt3 small interfering RNA (siRNA) into newly hatched larvae were well tolerated and led to a transient but significant decrease in the amount of Dnmt3 message, whereas this treatment was lethal for embryos (*21*). Both quantitative polymerase chain reaction (PCR) and in situ hybridization show that the amount of Dnmt3 mRNA is lower in siRNA-treated larvae than in control individuals (Fig. 1), with the strongest silencing occurring 48 to 50 hours post-injection, a time that coincides with the L2-to-L3 larval transition, a critical "decision-making" period in larval development (*7, 11*). In contrast, injections of a "nonlarval" gene, *uth* (*22*), had no detectable effects on the amount of Dnmt3 mRNA.

This interference with Dnmt3 expression triggers profound developmental changes leading to contrasting adult outcomes (Fig. 2; see fig. S1 for data from individual experiments). In the Dnmt3 siRNA-treated larvae, the majority of emerging adults (72%) were queens with fully developed ovaries, and the remaining 28% were typical workers with rudimentary ovaries (Fig. 2). In contrast, in the larval group injected with a control gene siRNA (*uth*), 77% of adults emerged as workers, whereas the remaining 23% exhibited queenlike morphological features. However, these queenlike individuals had grossly underdeveloped ovaries with no more than 50 to 80 ovarioles per ovary compared with at least 120 to 190 ovarioles in those queens emerging in the Dnmt3 siRNA experiment (Fig. 2). The ovaries of siRNA-induced queens are practically indistinguishable from the ovaries of a virgin queen reared in the hive on pure royal jelly. Workers have only rudimentary ovaries with two to six ovarioles.

Do these phenotypic effects correlate with methylation changes in larval DNA? Because global methylation in the honeybee genome is low, we examined the methylation status of cytosines in a single gene, *dynactin p62*, that we had previously shown to be differentially methylated during development [figure S3 in (*19*)]. Dietary changes lead to differential expression of this gene in *Drosophila*, further underscoring its role in growth and feeding-related processes (*23, 24*).

As with all genes so far examined for CpG methylation in *Apis*, *dynactin p62* is methylated exclusively within the coding exonic sequences. Within the exonic landscape, we chose exons 5, 6, and 7 for bisulfite conversion because they contain a total of 10 CpG dinucleotides (Fig. 3), a high concentration compared with other genes we have examined. With this 0.5-kb bisulfite-converted and PCR-amplified fragment, we first mapped differences in *dynactin p62* methylation by using DNAs isolated from hive-reared whole larvae (late L3) destined to become either queens or workers. As shown in Fig. 3A, there is a detectable decrease in the overall amount of methylation of *dynactin p62* in the queen larvae (48%) versus the worker larvae (58%), suggesting that the methylation state of certain genes may correlate with the larva's developmental trajectory. However, because there are virtually no cell divisions during larval growth and larval tissues are both highly polyploid and heterogeneous with regard to their ploidy level (see examples in fig. S2), varying DNA dosages at any locus introduces technical limitations on methylation measurements in whole larvae. Therefore, to better evaluate the effectiveness of Dnmt3 RNAi silencing, we used only DNAs extracted from larval heads (late L3). Although the larval head contains several cell types (*25*),

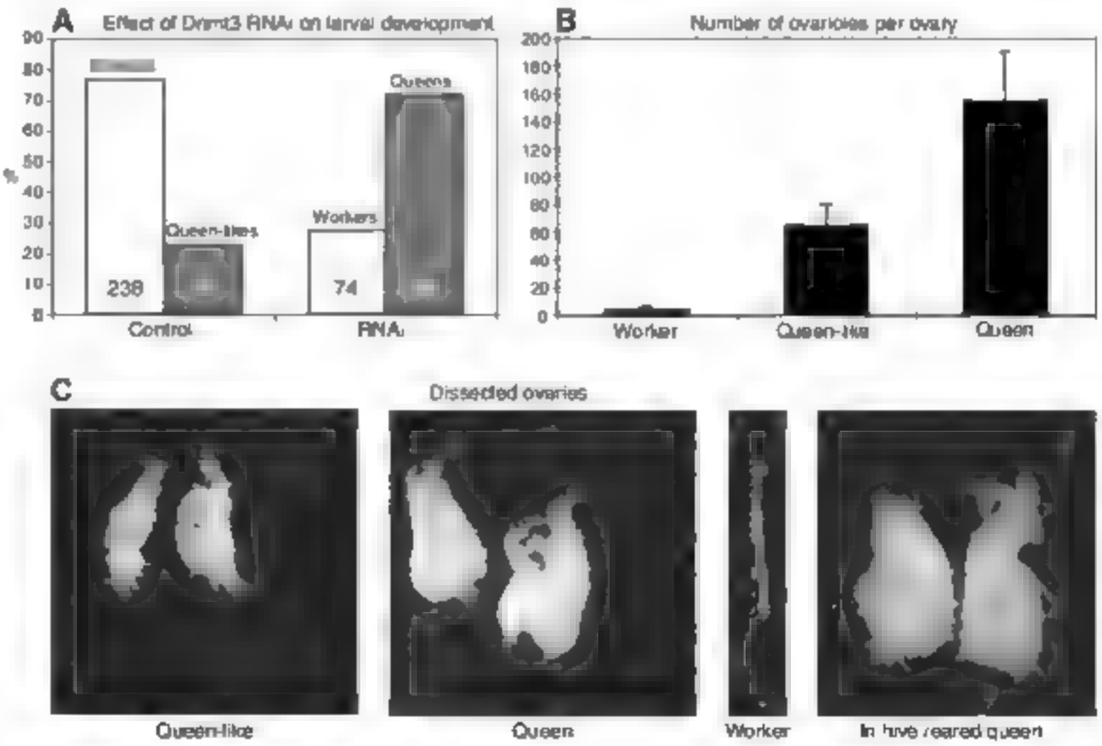


Fig. 2. Effect of Dnmt3 silencing on caste development in honeybees. Newly emerged larvae were injected either with a nonlarval control gene, *uth*, siRNA or with *Dnmt3* siRNA and allowed to develop until adulthood in a climate-controlled incubator. In both groups, the larvae developed normally, but the emerging adults displayed contrasting phenotypes. (**A**) The number of adults in each phenotypic category (workers, queens, and queenlikes). (**B**) The number of ovarioles per single ovary in each phenotypic class. Range error bars encompass the lowest and highest values. (**C**) Examples of ovaries dissected from each category and, for comparison, from a virgin queen reared in the hive on royal jelly. Queenlikes have queen morphological features but fewer ovarioles per ovary than queens [see (*8*)]. Workers have only rudimentary ovaries with two to six ovarioles. The figure is a compilation of four independent experiments. See (*21*), table S2, and fig. S1 for more details and results from individual experiments.

the tissues most relevant to developmental processes are brain neurosecretory cells and corpora allata (CA), the gland producing juvenile hormone implicated in the control of development and caste determination (*7*). The CA undergo several phases of endomitosis that lead to a high degree of its polyploidization in both queens and workers, but the rate of growth of the CA in queens is significantly accelerated, and at the completion of larval development (L5) the queen CA are twice the size of the worker CA (*26*). The high amount of Dnmt3 expression in CA shown in Fig. 1B may well be indicative of a link between the gland's function and its methylation status. Furthermore, the highly polyploid nature of the CA at stages L3 to L5 (fig. S2) must correlate with massive DNA replication that provides an opportunity for either adding or removing the methyl tags to target loci.

The results shown in Fig. 3 are consistent with this expectation. As illustrated in Fig. 3B, the 10% decrease in *dynactin p62* methylation in the heads of queen larvae is basically the same as the decrease found in the whole bodies, but the overall amount of CpG methylation in clones sequenced from both worker and queen larval heads is significantly higher than those analyzed from the whole bodies. The heads of worker larvae show 73% methylation across the *dynactin* fragment compared with only 58% in clones extracted from the workers' whole bodies, and for the queen larvae the head clones show 63% methylation versus 48% in the whole bodies (Fig. 3, A and B, right). This finding suggests that during larval development a high amount of CpG methylation and/or demethylation is associated with selected tissues, most likely with those that undergo massive DNA replication, such as cells of the CA and neurosecretory cells.

The analysis of the Dnmt3 RNAi silencing experiments in laboratory-reared larvae reveals a profile that is remarkably similar to that seen in the hive-reared individuals (Fig. 3C). The heads of worker larvae in the control group (worker-destined) show 79% methylation across the *dynactin p62* fragment, whereas the clones extracted from the Dnmt3 siRNA-injected larvae (queen-destined) show 63% methylation (Fig. 3C).

Interestingly, individual CpG sites reveal greater differences between the castes than those illustrated by the average methylation amounts across the entire *dynactin* fragment. For CpG sites 2 and 4, the decrease in cytosine methylation between the control and siRNA-treated clones is more than 30%, and for CpG site 10 the difference is 25% (Fig. 3C). These CpGs also show more than the average differential methylation in the heads of hive-reared larvae (Fig. 3B). This finding suggests that certain CpG sites might be preferentially methylated, but further studies are required to determine whether CpG methylation in *Apis* is used for transcriptional silencing of individual genes or is part of a global mechanism controlling transcriptional domains across the whole genome.

To identify members of networks regulated by methylation during larval growth and to gain some understanding of the epigenetic hierarchy that leads to alternative developmental paths we used the honeybee genomic oligonucleotide microarray to compare global gene expression between the control and Dnmt3-silenced lar-

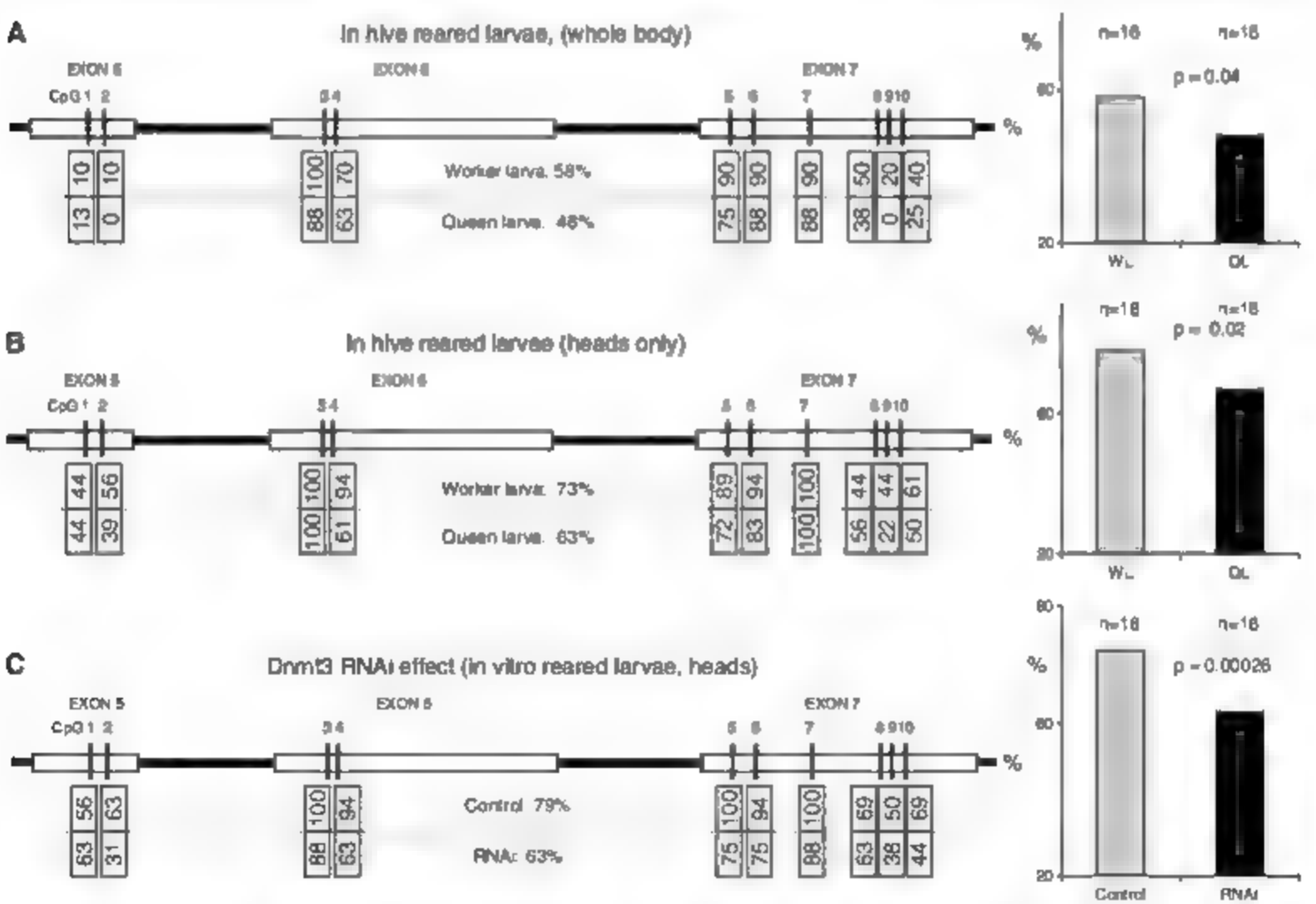


Fig. 3. Methylation status of cytosines in CpG dinucleotides of *dynactin p62*. The percentage of methylation for individual CpGs is shown in boxes, and the overall methylation in the right-hand graphs. DNA was isolated by using larvae collected (**A** and **B**) from the hive [for (A), pooled whole late-L3 larvae, $n = 7$; for (B), heads only, $n = 20$ for workers and $n = 14$ for queen larvae] and (**C**) from pooled heads of late-L3 in vitro reared larvae ($n = 7$). The number of clones sequenced for each category is shown above the bars in the right-hand graphs. Methylation quantities along this gene were analyzed with a general linear model of the binomial family (*32*) by using treatment (diet or RNAi) and position as factors to model the state of each CpG. The differences between queen larvae (QL) and worker larvae (WL) as well as the effect of RNAi are statistically significant.

vae (table S1, ArrayExpress accession number E-MEXP-1394). This revealed a battery of differentially expressed genes associated with lipid transport, hormonal regulation, posttranslational modification, protein turnover, ribosomal biogenesis, energy transfer, and other physiometabolic processes, as well as a number of novel, possibly *Apis*-specific genes with unknown functions (table S1). One differentially expressed gene predicted to encode an ortholog of the adenosine triphosphatase (ATPase) Belphegor has been shown in *Drosophila* to be regulated by TOR, a serine/threonine kinase that is central to a signaling cascade controlling growth (*27*). TOR is believed to be at the core of an ancient gene network that senses nutrient levels, and its involvement in honeybee queen development has already been considered (*7*, *28*). We also expected that the set of genes responsive to Dnmt3 silencing might contain genes implicated in chromatin remodeling and integrity. It is known that regulation of eukaryotic gene expression requires two classes of chromatin-remodeling enzymes: those that modify histones through acetylation, phosphorylation, or methylation and those that alter chromatin structure through hydrolysis of adenosine triphosphate (ATP) (*29*). From the list shown in table S1, we identified at least two genes belonging to this category, a subunit of the INO80 nucleosome remodeling complex and an ATPase of the type involved in structural maintenance of chromosomes. The transcriptional responses to Dnmt3 silencing are in good agreement with our previous study in which we used a smaller, expressed sequence tag (EST) based array to compare gene expression between queen and worker larvae extracted from the colony environment (*7*). Thus, both Dnmt3 silencing and feeding with royal jelly induce reprogramming of the larval transcriptome that is characterized by transcriptional shift toward higher expression of physiometabolic genes, including genes coding for metabolic enzymes and the general growth of the organism.

This study shows that in *A. mellifera* DNA methylation is a key component of an epigenetic network controlling a most important aspect of eusociality, the reproductive division of labor (*30*). Further work is required to unravel the causal relation between diet-induced methylation changes and altered gene expression, but our data hold substantial promise for functional methylome analyses in Metazoa using an easily manipulatable insect system. Where appropriate, the relevance of the honeybee findings to mammals can be evaluated to determine the level of data transferability, especially in the context of nutritional processes, longevity, and even drug treatment. Methylation data on a genome scale combined with genomewide expressional profiling in both social and solitary insects possessing functional CpG methylation systems will be needed to dissect the intricacies of this elaborate regulatory system.

References and Notes

1. M. West-Eberhard, *Annu. Rev. Ecol. Syst.* **20**, 249 (1989).
2. M. J. West-Eberhard, *Developmental Plasticity and Evolution* (Oxford Univ. Press, New York, 2003).
3. E. O. Wilson, *Proc. Am. Philos. Soc.* **123**, 204 (1979).
4. R. Hartfelder, W. Engels, *Curr. Top. Dev. Biol.* **40**, 45 (1998).
5. J. D. Evans, D. E. Wheeler, *Proc. Natl. Acad. Sci. U.S.A.* **96**, 5575 (1999).
6. J. D. Evans, D. E. Wheeler, *Genome Biol.* **2**, 1 (2000).
7. A. R. Barchuk *et al.*, *BMC Dev. Biol.* **7**, 70 (2007).
8. E. H. Calhoun, M. V. Smith, *Nature* **188**, 854 (1960).
9. R. W. Shuel, S. E. Dixon, *Insectes Soc.* **7**, 265 (1960).
10. N. Weaver, *Ann. Entomol. Soc. Am.* **50**, 283 (1957).
11. N. Weaver, *J. Econ. Entomol.* **50**, 759 (1957).
12. R. Jaenisch, A. Bird, *Nat. Genet.* **33** (suppl.), 245 (2003).
13. G. C. Burdge, M. A. Hanson, J. L. Slater-Jefferies, K. A. Lillycrop, *Br. J. Nutr.* **97**, 1036 (2007).
14. G. Kaati, L. O. Bygren, S. Edvinsson, *Eur. J. Hum. Genet.* **10**, 682 (2002).
15. R. A. Waterland, R. L. Jirtle, *Mol. Cell. Biol.* **23**, 5293 (2003).
16. V. K. Rakyan, M. E. Blewitt, R. Druker, J. L. Preis, E. Whitelaw, *Trends Genet.* **18**, 348 (2002).
17. I. C. Weaver *et al.*, *Nat. Neurosci.* **7**, 847 (2004).
18. M. G. Goll, T. H. Bestor, *Annu. Rev. Biochem.* **74**, 481 (2005).
19. Y. Wang *et al.*, *Science* **314**, 645 (2006).
20. M. Schaefer, F. Lyko, *Bioessays* **29**, 208 (2007).
21. Materials and methods are available on *Science* Online.
22. J. Maleszka, S. Forêt, R. Saint, R. Maleszka, *Dev. Genes Evol.* **217**, 189 (2007).
23. FlyAtlas. http://flyatlas.org/atlas.cgi?name=FBgn0033206.
24. GEO (Gene Expression Omnibus). www.ncbi.nlm.nih.gov/sites/entrez?db=geo&cmd=search&term=FBgn0033206.
25. J. A. Nelson, *The Embryology of the Honeybee* (Princeton Univ. Press, Princeton, NJ, 1915).
26. G. M. Ulrich, H. Rembold, *Cell Tissue Res.* **230**, 49 (1983).
27. D. A. Guertin, K. V. Guntur, G. W. Bell, C. C. Thoreen, D. M. Sabatini, *Curr. Biol.* **16**, 958 (2006).
28. A. Patel *et al.*, *PLoS ONE* **2**, e509 (2007).
29. C. J. Fry, C. L. Peterson, *Science* **295**, 1847 (2002).
30. M. D. Drapeau, S. Albert, R. Kucharski, C. Prusko, R. Maleszka, *Genome Res.* **16**, 1385 (2006).
31. J. Maindonald, J. Braun, *Data Analysis and Graphics Using R* (Cambridge Univ. Press, Cambridge, 2007).

32. We thank G. Milkos for stimulating discussion and comments on the manuscript and P. Hellwell for beekeeping and excellent technical assistance. This work was supported by the Australian Research Council Discovery grant DP0770821 awarded to R.M. GenBank accession no. for *dynactin p62* is XP_001121083.

Supporting Online Material

www.sciencemag.org/cgi/content/full/1153069/DC1
Materials and Methods
Figs. S1 to S4
Tables S1 and S2
References

16 November 2007; accepted 27 February 2008
Published online 13 March 2008.
10.1126/science.1153069
Include this information when citing this paper.

The Flavivirus Precursor Membrane-Envelope Protein Complex: Structure and Maturation

Long Li, Shee-Mei Lok, I-Mei Yu, Ying Zhang, Richard J. Kuhn, Jue Chen, Michael G. Rossmann*

Many viruses go through a maturation step in the final stages of assembly before being transmitted to another host. The maturation process of flaviviruses is directed by the proteolytic cleavage of the precursor membrane protein (prM), turning inert virus into infectious particles. We have determined the 2.2 angstrom resolution crystal structure of a recombinant protein in which the dengue virus prM is linked to the envelope glycoprotein E. The structure represents the prM-E heterodimer and fits well into the cryo–electron microscopy density of immature virus at neutral pH. The pr peptide β-barrel structure covers the fusion loop in E, preventing fusion with host cell membranes. The structure provides a basis for identifying the stages of its pH-directed conformational metamorphosis during maturation, ending with release of pr when budding from the host.

Many viruses, including flaviviruses (*1*), undergo a maturation step immediately before their release from the host. The evident purpose of this step is to maintain stability for the hazardous transfer to a new host while preparing virions for rapid fusion with, and entry into, a cell. Flaviviruses within the *Flaviviridae* family are major human pathogens that include dengue virus, West Nile virus, yellow fever virus, and Japanese encephalitis virus. They have a positive-sense, 11-kb RNA genome that is packaged together with multiple copies of the capsid protein within a lipid envelope (*2*). The genome is translated as a polyprotein that has the capsid protein, the precursor membrane glycoprotein (prM), and the envelope glycoprotein (E) in its N-terminal region (Fig. 1A). The polyprotein is cleaved into component proteins by viral and cellular proteases (*2*). Partially assembled flavivirus nucleocapsids bud from the endoplasmic reticulum, thereby becoming enveloped with a lipid membrane that comes with it the E and prM glycoproteins (*2*). These immature particles are transported through the cellular secretory pathway, where the cellular furin protease cleaves prM, eventually resulting in the release of the pr peptide and formation of mature virions (*3*, *4*).

The dengue virus prM glycoprotein consists of 166 amino acids. Cleavage by furin releases

Department of Biological Sciences, Purdue University, West Lafayette, IN 47907, USA.

*To whom correspondence should be addressed. E-mail: mr@purdue.edu

the N-terminal 91 "pr" residues during maturation, leaving only the ectodomain (residues 92 to 130) and C-terminal transmembrane region (residues 131 to 166) of "M" in the virion. The pr peptide protects immature virions against premature fusion with the host membrane (*5*, *6*). The dengue virus E glycoprotein participates in the fusion of the virion with the endosomal membrane at low pH. It consists of an ectodomain (soluble E protein, sE), a stem region, and a transmembrane domain. The x-ray crystallographic structure of sE has been determined for a number of flaviviruses (*7*–*12*), all of which have three domains (DI, DII, and DIII) that consist mainly of β sheets with the fusion loop at the distal end of DII. The E protein is able to switch among different oligomeric states: as a trimer of prM-E heterodimers in immature particles, as a dimer in mature virus, and as a trimer when fusing with a host cell (*8*, *10*).

The cryo–electron microscopy (cryoEM) structures of immature flaviviruses have been determined at neutral pH (*6*, *13*). The "spiky" icosahedral immature virions have a diameter of about 600 Å and contain 60 trimeric prM-E spikes. In contrast, the final "smooth"-surfaced icosahedral mature particles have a diameter of about 500 Å and contain 90 E dimers arranged in a herringbone pattern and 180 copies of the M protein (*14*, *15*). The transformation from immature to mature particles requires some large rearrangement of the E and M proteins (Fig. 2) (*12*, *13*).

A recombinant fusion protein of prM and E from dengue virus 2 was constructed in which the transmembrane region of prM was replaced with an 8–amino acid linker (Fig. 1B) (*16*). The furin cleavage site of prM was mutated to prevent cleavage of the recombinant protein by intracellular proteases. The crystal structure of the recombinant protein was determined at pH 5.5 to 2.20 Å resolution, and also at pH 7.0 to 2.60 Å resolution (table S1). There were no significant structural differences between the two determinations, which had a root mean square difference of 1.0 Å between all pairs of equivalent atoms. Because of the slightly higher resolution of the low-pH structure, it was chosen for all subsequent calculations and discussions. The polypeptide chain of much of the prM protein (residues 1 to 81, corresponding to most of the pr peptide) and most of the sE protein could be traced in the electron density of the prM-E crystal structure. The pr peptide was positioned over the fusion loop at the distal end of DII (Fig. 3A), as anticipated given that it functions to prevent membrane fusion (*5*).

The pr peptide consists of seven β strands that are mostly antiparallel (Fig. 3). Three disulfide bonds (C34-C68, C45-C80, and C53-C66) stabilize the pr peptide structure, and the electron density map shows that Asn^{69} is glycosylated. A DALI search (*17*) did not find any structures with significant similarity to that of the pr peptide. The structure of the E protein in

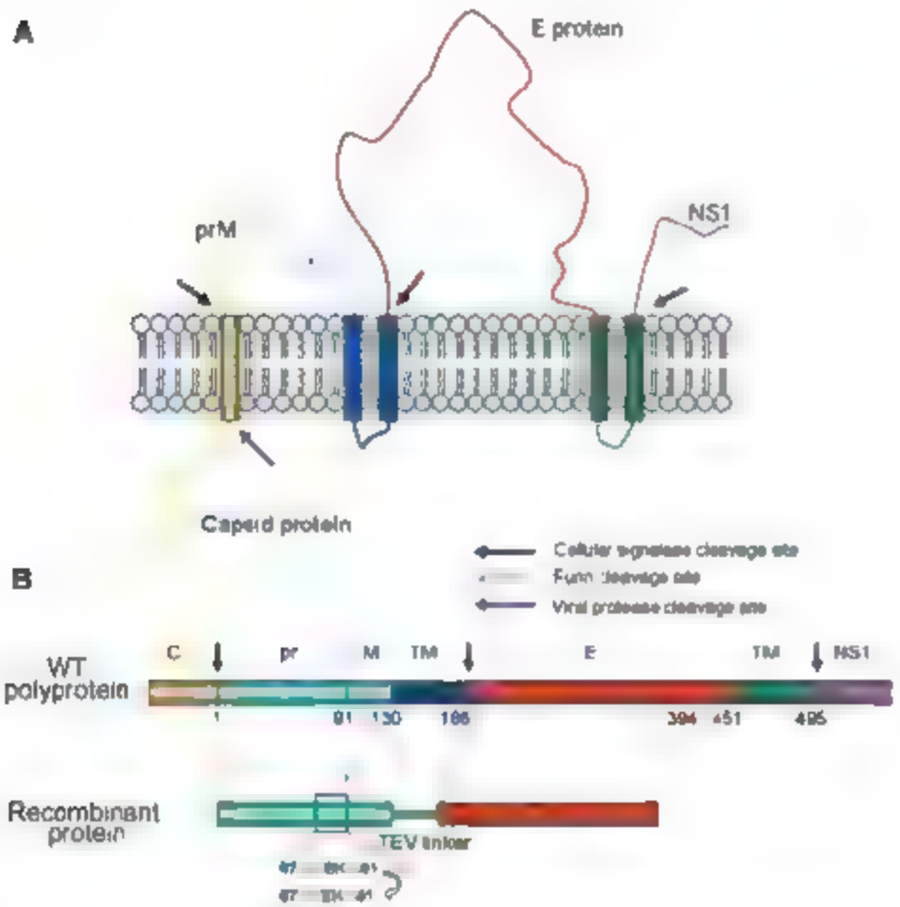


Fig. 1. The dengue virus polyprotein and the recombinant protein containing the prM and E proteins. (**A**) Threading of the dengue virus polyprotein N-terminal region through the endoplasmic reticulum membrane, showing the positions of the capsid, prM, and E structural proteins. Different-colored arrows indicate various protease cleavage sites. (**B**) The order of viral proteins in the wild-type polyprotein and in the recombinant protein construct. The proteins are identified by the same colors as in (A). The mutations to inhibit furin cleavage in the recombinant protein are shown below (*21*). The linker between the prM and E proteins is labeled TEV.

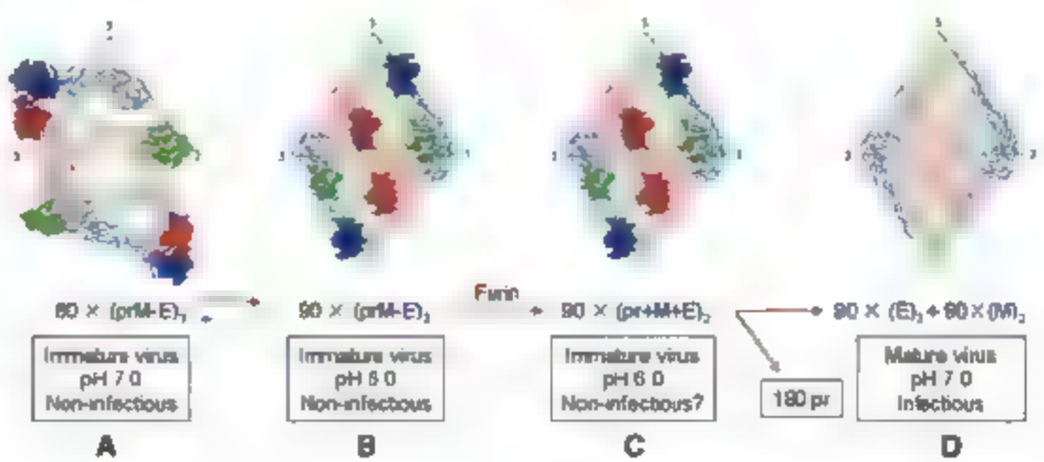


Fig. 2. Rearrangement of the prM and E proteins during virus maturation. (**A** to **D**) Sequence of events as referenced in the text. The E proteins are shown as a Cα backbone; space-filling atoms show the pr peptide surfaces. The three independent heterodimers per icosahedral asymmetric unit are colored red, green, and blue. Although the diagram assumes knowledge of the relationship among the positions of specific heterodimers in the immature and mature viruses (red goes to red, green to green, and blue to blue), this is not known.

the prM-E heterodimer is similar to the crystal structure of the E protein in the dimeric, prefusion form (*12*). The hinge angle between DI and DII is only 5° different from the structure of E in immature virus (*12*), as compared to 23° with the mature virus, which suggests that the oligomeric state of the E protein determines the hinge angle. This similarity supports the biological relevance of the recombinant fusion protein structure. The contact area between pr and E is 865 Å^2, representing 16% of the surface area of pr and 4% of E. There are three prominent complementary electrostatic patches (Fig. 3C) involving (Arg^6), (Glu^{46}, Asp^{47}), and (Asp^{63}, Asp^{65}) on pr, and involving (Glu^{64}), (Lys^{64}), and (His^{244}, Lys^{247}) on E (table S2), respectively. Of these, the pr residues Asp^{63} and Asp^{65} and the E residue His^{244} are strictly conserved among all known flavivirus sequences.

A pseudo-atomic structure of the immature dengue virus at neutral pH (fig. S1A) was generated by fitting the prM-E crystal structure into the 12.5 Å resolution cryoEM density map (*12*). The structure of the pr peptide fits the density well, including the prominent carbohydrate moieties at residue Asn^{69} (fig. S1B and table S3). The surface area buried between pairs of heterodimers is 1052 Å^2, 1445 Å^2, and 0 Å^2 in the "blue-red," "red-green," and "green-blue" interfaces, respectively, showing nonequivalent contacts between each of the three pairs of heterodimers (Fig. 4 and fig. S1). The hydrophobic fusion loop in each of the three E proteins within one spike is covered and surrounded by loops of the pr peptides and the carbohydrate moieties associated with Asn^{69} in pr, thus making the surface of the immature particle more hydrophilic (Fig. 4B) and protecting the E protein from premature fusion.

The cryoEM density representing the surface spikes was set to zero at all points within 3.0 Å of every atom in the fitted x-ray structure. There then remained traces of density that ran along the edge of each E molecule toward the lipid membrane (fig. S1C and fig. S2, A to C). These density traces were positioned similarly on each E molecule, in agreement with density in the crystal structure that had been presumed to be the N-terminal region of the M protein (fig. S2D) (*16*). The trace of the M protein suggested that the pr polypeptide chain is extended linearly along the surface of the E protein, mostly on the inside of the spike (fig. S1C). The position of the furin cleavage site could be reasonably well inferred by building an extended polypeptide chain into the density traces representing the junction of the pr peptide with the M protein (Fig. 4C and movie S1). Docking of the known structure of furin (*18*) onto the cleavage site showed that furin would be sterically hindered from binding to any of the three prM-E heterodimers within a spike, thereby demonstrating why furin is unable to

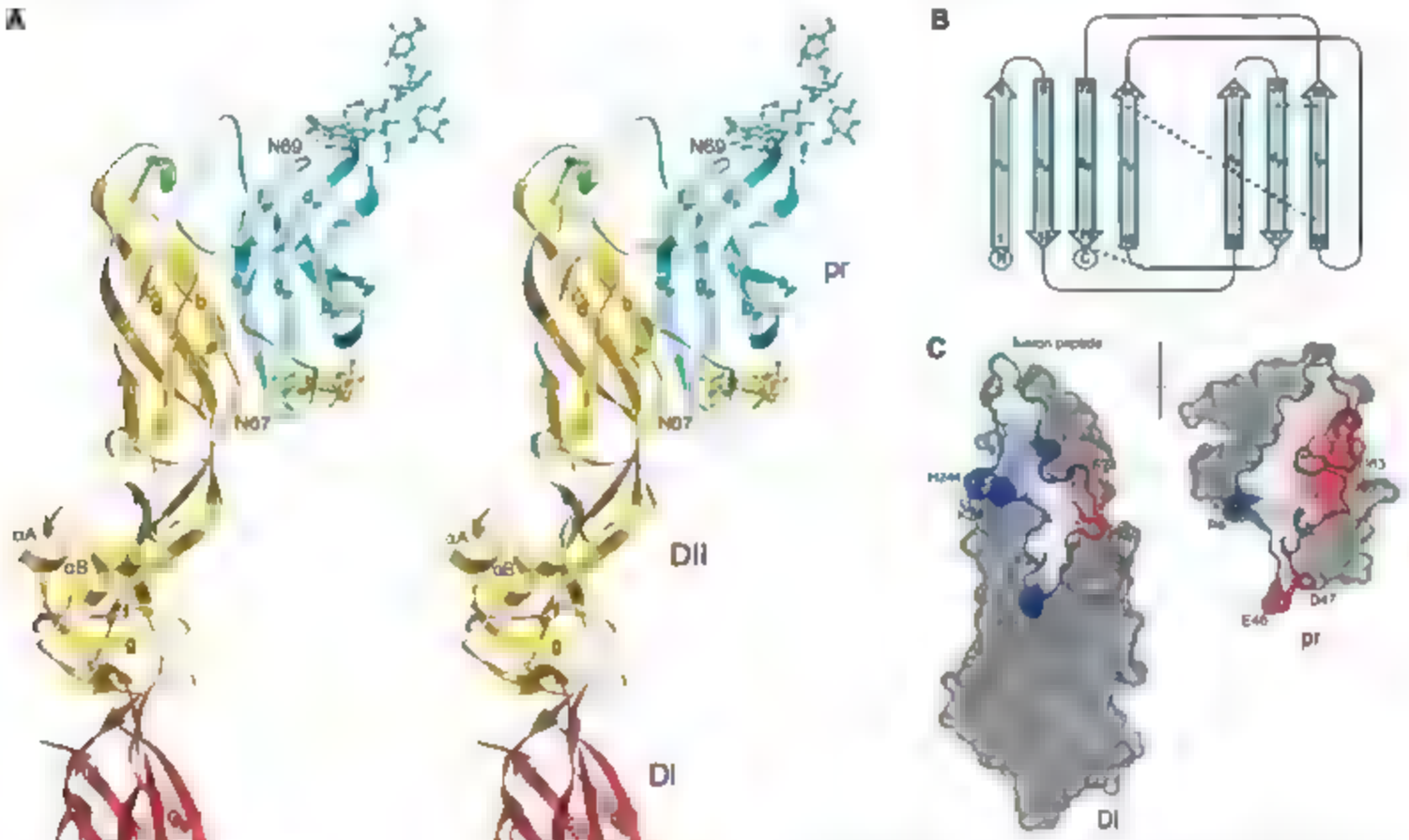


Fig. 3. Structure of the prM-E protein. (**A**) Stereoview of the prM-E heterodimer in ribbon representation. The pr peptide is cyan, DI is red, DII is yellow, DIII is blue, and the fusion loop is green. Secondary structural elements of the E protein are labeled (*7*, *21*). Secondary structural elements for the pr peptide are defined here as a_{pr}, b_{pr}, ..., g_{pr}. (**B**) Secondary structure of the pr peptide. Disulfide linkages are indicated with dashed lines. (**C**) Open-book view of the pr-E interactions showing charge complementarity. Positively and negatively charged surfaces in the contact areas are colored blue and red, respectively. Charged residues in the contact area are labeled. The fusion loop (residues 100 to 108) is outlined in green. Contact areas are defined by atoms less than 4.5 Å apart between the pr peptide and the E protein (*21*).

cleave the pr polypeptide in the immature virus at neutral pH (*4*).

The accompanying paper (*4*) shows that low-pH immature virus particles (Fig. 2B) have a structure in which the arrangement of the E proteins is essentially the same as that of the mature virus (Fig. 2D). However, the interface between the DII domains in the E dimer is in part the same surface where the extended polypeptide of the M protein binds in the neutral-pH immature virus (fig. S3A). Thus, if the M protein were in the conformation as found in the immature virus at neutral pH, it would sterically block the formation of the E dimer. Hence, the conformation of M must be different in the low-pH immature virus, consistent with its apparent flexibility in the crystal structure of the heterodimer and also consistent with the changed position of the M protein's transmembrane helices during maturation (fig. S3) (*15*). Indeed the low-pH conformation, unlike the neutral-pH conformation of the M protein, has been found to be accessible to furin cleavage (*4*).

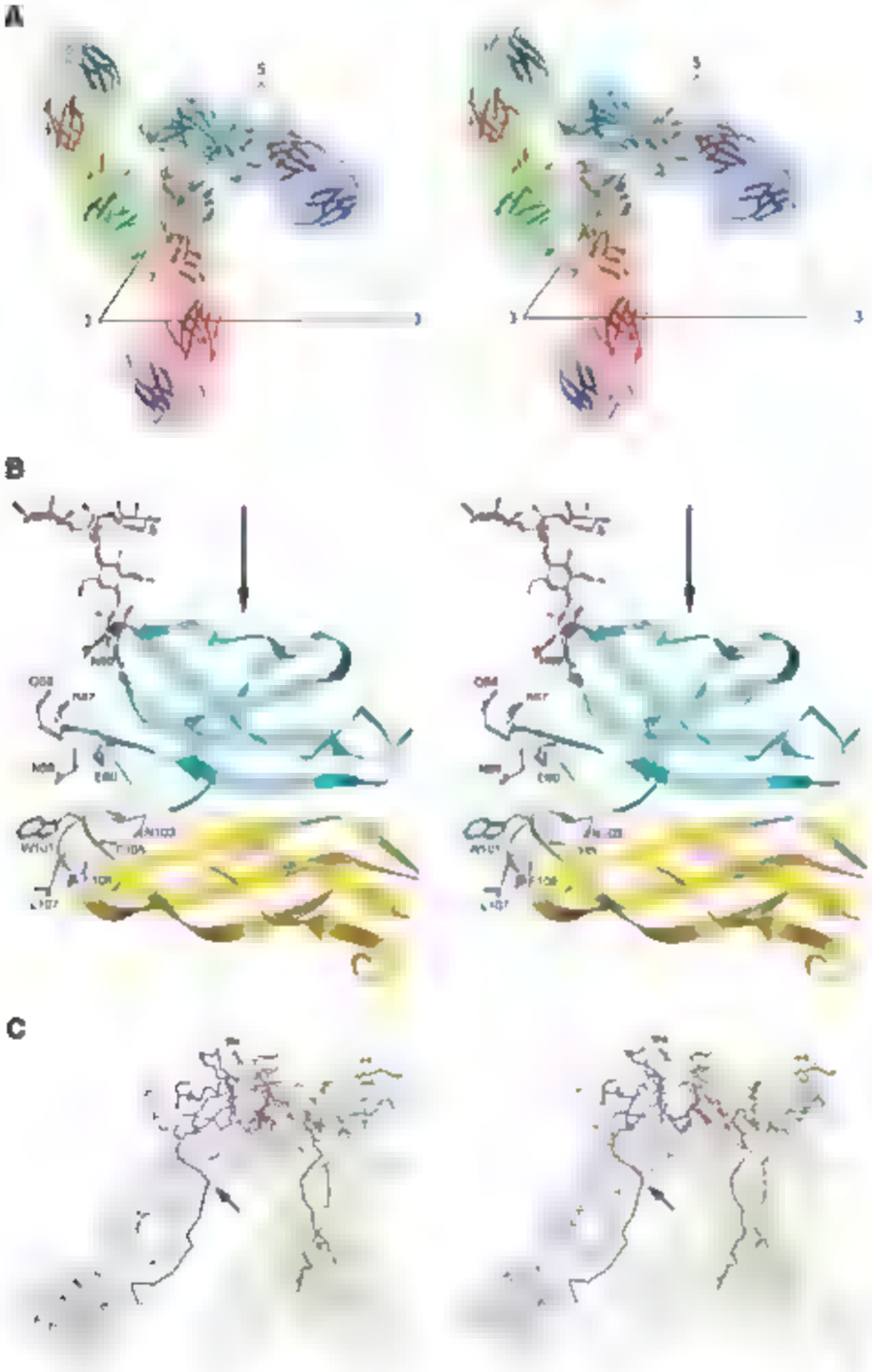


Fig. 4. Pseudo-atomic structure of the neutral-pH immature dengue virus. (**A**) Stereoview of one spike of the immature virus as seen from outside the virus, colored as in Fig. 3A. The background to each heterodimer is colored red, green, or blue in accordance with the color code used in Fig. 2 and table S3. (**B**) Stereoview showing the fusion loop of the E molecule and the protecting pr loops and glycans (*21*). The arrow indicates the direction of viewing from outside the virus. (**C**) Stereoview of the fitted pr-E Cα backbones with the Cα trace of the extended M polypeptide (black) running along the edge of each E protein. The Cα backbones of the pr peptides are in bold. The approximate site of furin cleavage is marked with a black arrow in the blue molecule.

The large conformational change that occurs when the immature virus changes from the neutral-pH to the low-pH form was found to be reversible for dengue virus (Fig. 2, A and B), as long as prM was still intact (*4*). However, once prM had been cleaved (Fig. 2C), there was no further conformational change when the pH was returned to neutral (Fig. 2, C and D), instead, the cleaved pr peptide was released (*4*). Apparently the extended polypeptide of the M protein, along the side of the E protein (Fig. 4C and fig. S2), is essential for maintaining the reversibility of the conformational change (Fig. 2, A and B). An analogy might be the effect of a drawstring that opens and closes a curtain. Once the string (i.e., the M protein) is cut (i.e., furin cleavage), there can be no further movement. Mutating a conserved histidine residue to alanine in the M protein ($His^{98} \rightarrow Ala$) in Japanese encephalitis virus inhibits the formation of prM-E heterodimers (*19*). The corresponding residue, His^{98} in dengue virus, is located approximately in the center of the extended M protein, opposite the hydrophobic surfaces of helices αA and αB in the E protein. Thus, a change of pH might alter the interactions between M and E, leading to the transition between the "spiky" and "smooth" virus conformations.

On one hand, pr remains bound to E when the immature low-pH virus is returned to neutral pH, thus protecting the immature virus against fusion. On the other hand, after cleavage of M, pr is released from E to activate the virus when returning the pH to neutral. The average area of contact between a pr peptide and an E protein in a spike of the immature virus at neutral pH is slightly larger than the area of contact in the immature virus at low pH (table S4). Thus, not only is each pr peptide tethered covalently to the M protein, but also the pr peptide probably has a slightly greater affinity for the trimeric E protein spike of immature virus at neutral pH relative to the dimeric smooth surface of mature virus. In contrast, His^{244} of the E protein is highly conserved and is situated in the prM-E interface opposite the completely conserved Asp^{63} of the pr peptide (Fig. 3C), causing the affinity of the pr peptide for the E protein to decrease when the pH is raised to neutral. This would allow the pr peptide to be released at neutral pH from the dimeric E structure, but only when the pr peptide has been cleaved. As evolution is highly conservative of structure (*20*), the maturation process described here for dengue virus is likely to have structural homologies in other enveloped viruses.

References and Notes

1. S. Mukhopadhyay, R. J. Kuhn, M. G. Rossmann, *Nat. Rev. Microbiol.* **3**, 13 (2005).
2. B. D. Lindenbach, C. M. Rice, in *Fields Virology*, D. M. Knipe, P. M. Howley, Eds. (Lippincott Williams & Wilkins, Philadelphia, 2001), pp. 991–1041.
3. K. Stadler, S. L. Allison, J. Schalich, F. X. Heinz, *J. Virol.* **71**, 8475 (1997).
4. I.-M. Yu *et al.*, *Science* **319**, 1834 (2008).
5. F. Guirakhoo, R. A. Bolin, J. T. Roehrig, *Virology* **191**, 921 (1992).
6. Y. Zhang *et al.*, *EMBO J.* **22**, 2604 (2003).
7. F. A. Rey, F. X. Heinz, C. Mandl, C. Kunz, S. C. Harrison, *Nature* **375**, 291 (1995).
8. S. Bressanelli *et al.*, *EMBO J.* **23**, 728 (2004).
9. Y. Modis, S. Ogata, D. Clements, S. C. Harrison, *Proc. Natl. Acad. Sci. U.S.A.* **100**, 6986 (2003).
10. Y. Modis, S. Ogata, D. Clements, S. C. Harrison, *Nature* **427**, 313 (2004).
11. R. Kanai *et al.*, *J. Virol.* **80**, 11000 (2006).
12. Y. Zhang *et al.*, *Structure* **12**, 1607 (2004).
13. Y. Zhang, B. Kaufmann, P. R. Chipman, R. J. Kuhn, M. G. Rossmann, *J. Virol.* **81**, 6141 (2007).
14. R. J. Kuhn *et al.*, *Cell* **108**, 717 (2002).
15. W. Zhang *et al.*, *Nat. Struct. Biol.* **10**, 907 (2003).
16. See supporting material on *Science* Online.
17. L. Holm, C. Sander, *Science* **273**, 595 (1996).
18. S. Heinrich *et al.*, *Nat. Struct. Biol.* **10**, 520 [illegible].
19. Y.-L. Lin, S.-C. Wu, *J. Virol.* **79**, 8535 (2005).
20. M. G. Rossmann, D. Moras, K. W. Olsen, *Nature* **250**, 194 (1974).
21. Single-letter abbreviations for amino acid residues: A, Ala; C, Cys; D, Asp; E, Glu; F, Phe; G, Gly; H, His; I, Ile; K, Lys; L, Leu; M, Met; N, Asn; P, Pro; Q, Gln; R, Arg; S, Ser; T, Thr; V, Val; W, Trp; Y, Tyr.

22. We thank Y. Xiang and W. Zhang for many helpful discussions, S. Kelly and C. Towell for help in the preparation of the manuscript, and the staff at APS GM/CA sector for their help in data collection. The facilities are supported by the U.S. Department of Energy and/or NIH. The work was supported by NIH program project grant AI055672 (R.J.K., M.G.R., J.C.) and National Institute of Allergy and Infectious Diseases Region V Great Lakes Center of Excellence for Biodefense and Emerging Infectious Diseases Research Program award 1-U54-AI-057153 (R.J.K., M.G.R.). The coordinates of the prM-E heterodimer crystal structures at pH 5.5 and 7.0 have been deposited with the Protein Data Bank (accession numbers 3C5X and 3C6E, respectively). The fit of the prM-E heterodimer into the cryoEM reconstruction of the dengue immature virus at neutral pH has been deposited with the RCSB Protein Database (accession number 3C6D).

Supporting Online Material

www.sciencemag.org/cgi/content/full/319/5871/1834/DC1

Materials and Methods
Figs. S1 to S3
Tables S1 to S4
Movie S1
References

21 November 2007; accepted 29 February 2008
10.1126/science.1153263

Structure of the Immature Dengue Virus at Low pH Primes Proteolytic Maturation

I-Mei Yu, Wei Zhang, Heather A. Holdaway, Long Li, Victor A. Kostyuchenko, Paul R. Chipman, Richard J. Kuhn, Michael G. Rossmann, Jue Chen*

Intracellular cleavage of immature flaviviruses is a critical step in assembly that generates the membrane fusion potential of the E glycoprotein. With cryo–electron microscopy we show that the immature dengue particles undergo a reversible conformational change at low pH that renders them accessible to furin cleavage. At a pH of 6.0, the E proteins are arranged in a herringbone pattern with the pr peptides docked onto the fusion loops, a configuration similar to that of the mature virion. After cleavage, the dissociation of pr is pH-dependent, suggesting that in the acidic environment of the trans-Golgi network pr is retained on the virion to prevent membrane fusion. These results suggest a mechanism by which flaviviruses are processed and stabilized in the host cell secretory pathway.

The structure of viruses is dynamic because major conformational changes are necessary for the virus to enter and disassemble in a host cell. Whereas such conformational change is triggered by receptor binding or acidification, the origin of the labile conformation is generated in the assembly pathway. For example, glycoproteins of many enveloped viruses are synthesized as inactive precursors that require proteolytic cleavage to prime their membrane fusion potential. In the case of class I fusion proteins, represented by influenza virus hemagglutinin (HA), cleavage of the HA0 precursor generates a metastable form that undergoes a low-pH–induced conformational change to facilitate membrane fusion with a host cell (*1*). In contrast, class II fusion proteins, found in flaviviruses and alphaviruses, fold cotranslationally with an auxiliary protein. Proteolytic activation involves cleavage of the auxiliary protein that is thought to free the fusion protein for subsequent conformational changes necessary for membrane fusion (*2*–*4*).

In the endoplasmic reticulum (ER) of an infected cell, newly assembled immature flaviviruses contain heterodimers of the auxiliary precursor membrane (prM) protein and the envelope (E) protein (*5*). Furin (*6*), a cellular protease primarily located in the trans-Golgi network (TGN), cleaves prM to generate mature particles where the pr peptides are released and the E proteins form homodimers (*7*, *8*). Crystal structures of the E protein at neutral pH (*7*, *9*–*12*) show that each polypeptide chain contains three domains: the structurally central domain (DI), the dimerization domain containing the fusion loop (DII), and the carboxy-terminal immunoglobulin-like domain (DIII). In the postfusion conformation observed at the endosomal pH, the E dimers rearrange into homotrimers, of which the fusion loops and the C-terminal membrane anchors are located at the same end (*13*, *14*). Thus, the membrane fusion process of flaviviruses appears to require dissociation of the E dimers and formation of the postfusion trimers.

Acidotropic reagents such as NH_4Cl that raise the pH of the TGN prevent furin cleavage, resulting in immature particles containing uncleaved prM molecules (*15*–*17*). Whereas the mature virion has a smooth surface on which 90 E dimers form a closely packed protein shell with a herringbone pattern (*8*), at neutral pH the immature particles propagated in the presence of NH_4Cl contain 60 prominent spikes, each consisting of a trimer of prM-E heterodimers (*18*, *19*). To address whether these particles are physiologically relevant, we subjected dengue virus immature particles to furin cleavage at various pH values (*20*). In agreement with previous work (*6*), prM could be cleaved only under acidic conditions (pH of 5.0 to 6.0) (Fig. 1A). Whereas uncleaved immature viruses at both pH = 8.0 and pH = 6.0 are mildly infectious, in vitro furin cleavage led to a 1000-fold increase in infectivity (Fig. 1B). Because the optimum enzymatic activity of furin was observed around neutral pH (fig. S1), the pH dependence of the cleavage likely reflects the accessibility of the cleavage sites on the viral surface, suggesting that the conformation of the immature virion at low pH is distinct from that of the neutral pH form (*18*).

Cryo–electron microscopy (cryo-EM) and image reconstruction show that immature dengue particles undergo a reversible conformational change at low pH (Fig. 1C). As previously reported (*18*, *19*), at pH = 8.0 the particles have a spiky surface with 60 icosahedrally arranged protrusions. Incubation at pH = 6.0 resulted in particles with a much smoother surface. The size of these particles was around 530 Å in diameter, which is substantially smaller than that of the neutral-pH immature form (600 Å). When back-neutralized to pH = 7.5, the smooth particles at low pH were converted to the spiky form, indicating that the pH effect is reversible. The reversibility of the conformational change was further demonstrated by in vitro furin activation experiments (fig. S2). No furin cleavage or activation was observed at pH = 7.5 for samples exposed to low pH and then back-neutralized. However, when the pH was again lowered, the particles could be activated by furin (fig. S2A). These results are consistent with the observation that the neutral-pH structures of West Nile virus

Department of Biological Sciences, 915 West State Street, Purdue University, West Lafayette, IN 47907–2054, USA.

*To whom correspondence should be addressed. E-mail: chenjue@purdue.edu

immature particles propagated with and without prior low-pH exposure were identical (*19*).

The structure of low-pH immature particles, determined at 25 Å resolution, shows a multilayer architecture composed of the glycoproteins prM and E, the lipid bilayer, and the nucleocapsid (Fig. 2, A and B). The conformational difference of the immature particles at neutral and low pH was extensive, evidenced by the absence of similarities in the cryo-EM density outside the nucleocapsid core (fig. S3). In contrast, comparison of the low-pH immature particle to the cleaved, mature virion shows strong similarities (fig. S3). In the difference map calculated by subtracting the densities of the mature virus from that of the low-pH immature virus, the only significant densities (above 3 σ) are the three independent peaks within one asymmetric unit in the outmost region of the particle (Fig. 2C). Fitting the crystal structure of the prM-E heterodimer (*21*) into the cryo-EM map of the low-pH immature particles shows that the difference peaks correspond accurately to the pr peptide in the crystal structure. The arrangement of E proteins is essentially the same as in the mature virion (Fig. 2D and table S1). The three nearly parallel E dimers are apparent on the virion surface, with the fusion loop buried at the interface of the pr and E dimer. The two pr peptides associated with an E dimer are located at equivalent positions, making extensive interactions with the DII of one E monomer and the DI of a neighboring E monomer (Fig. 2, E and F). This configuration provides an explanation for why the immature viruses are stable at low pH (*6*) whereas the mature particles undergo membrane fusion (*16*). Apparently, the presence of pr stabilizes the metastable E dimer, preventing its dissociation and the subsequent formation of the postfusion homotrimers (*13*, *22*, *23*).

In contrast to the results reported here, previous studies of the tick-borne encephalitis (TBE) virus show that the low-pH–induced conformational change of the immature virus is irreversible (*6*). Given that the sequence identity of the E proteins among flaviviruses is only about 40%, it seems possible that the lateral interactions between the E dimers in the herringbone configuration are stronger in TBE virus than in dengue and West Nile viruses and thus that immature TBE viruses do not revert to the trimeric-spike configuration upon back-neutralization.

The requirement of low pH for furin cleavage suggests that the proteolytic processing of immature flaviviruses occurs in the acidic environment of the TGN, where furin is abundant (*24*). If so, what prevents the cleaved virions from undergoing membrane fusion within the TGN? One possible explanation is that the furin cleavage product, pr, remains associated with virions. To test this hypothesis, we cleaved immature particles by furin at low pH and then subjected them to sucrose gradient sedimentation at either pH = 5.5 or pH = 8.0 (Fig. 3). At low pH, a substantial amount of pr comigrated with the virion, whereas at pH = 8.0 all pr was found in the top fractions of the gradient. Thus, at acidic pH the dissociation rate of pr from the virus particle is slower than that at neutral pH. At pH = 5.5, about 30% of pr was observed in the soluble fraction during the 2-hour course of the experiment. It is likely that in vivo all pr molecules remain associated with the virus particle until it is released to the extracellular milieu, a process that takes a few minutes to complete.

Our results, interpreted in light of previous structural and functional data, suggest a complete scheme for the flavivirus maturation pathway (Fig. 4). Inside an infected cell, the newly synthesized viral RNA and proteins assemble on the ER membrane and bud into the lumen of the ER (*3*). The immature virus particles, containing trimeric prM-E spikes (*18*, *19*), are transported through the secretory pathway. Acidification in the TGN induces a global rearrangement of the glycoproteins, resulting in exposure of the furin cleavage site. The prM protein is cleaved by furin in the TGN, but the proteolytic product pr stays associated with the virion to prevent membrane fusion. Upon release to the extracellular milieu, pr dissociates at neutral pH, and the resulting mature virus undergoes membrane fusion in the next infection cycle.

Flaviviruses are among the large group of enveloped viruses that undergo membrane fusion at low pH. The budding sites of these viruses can be either the ER or the cell surface. Regardless of the budding site, the membrane-anchored fusion proteins are folded in the ER and transported through the cellular secretory pathway. How do the fusion proteins pass through the TGN without acid inactivation or inducing membrane fusion? Perhaps the answer lies in the energy landscape of the fusion proteins. The biosynthetic precursor of the fusion protein folds into the

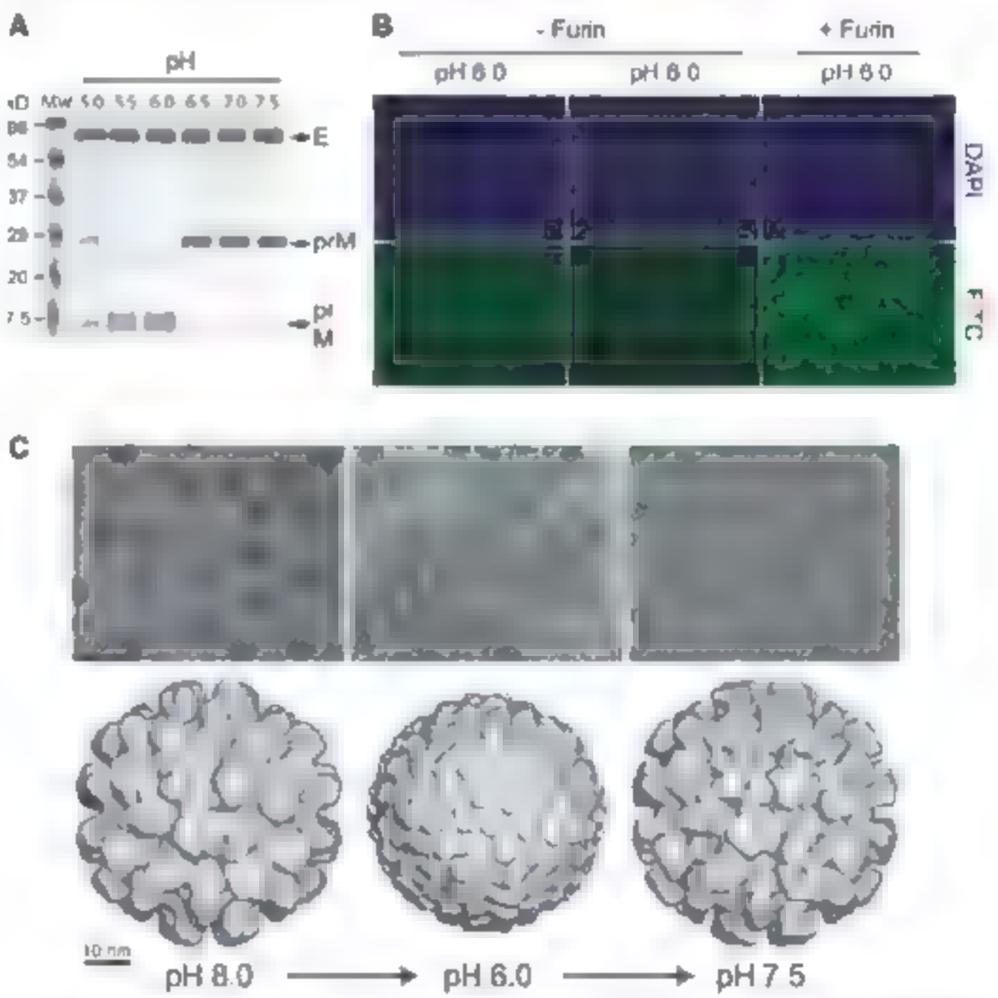


Fig. 1. A reversible conformational change induced by low pH. (**A**) Purified immature virus particles were incubated with furin at the indicated pH values and analyzed by nonreducing SDS–polyacrylamide gel electrophoresis (SDS-PAGE) followed by silver staining. It is likely that the pr peptide (91 residues), containing three disulfide bonds, migrates faster than the predicted molecular weight and thus was not resolved from the M protein (75 residues). (**B**) The infectivity of immature dengue particles with and without furin cleavage was analyzed by an immunofluorescence assay. Virus-infected BHK cells were probed with a monoclonal α-E antibody and detected by using a fluorescein isothiocyanate (FITC)–conjugated secondary antibody. Nuclei of all cells were counterstained with 4′,6′-diamidino-2-phenylindole (DAPI). The three samples shown here were infected with the same amount of virus particles. The amount of cells stained by FITC in the right image (+ furin) is about 1000-fold of those in the other two images (– furin). (**C**) Electron micrographs and shaded-surface representations of the immature particles at pH = 8.0 (left), pH = 6.0 (middle), and back-neutralized to pH = 7.5 (right).

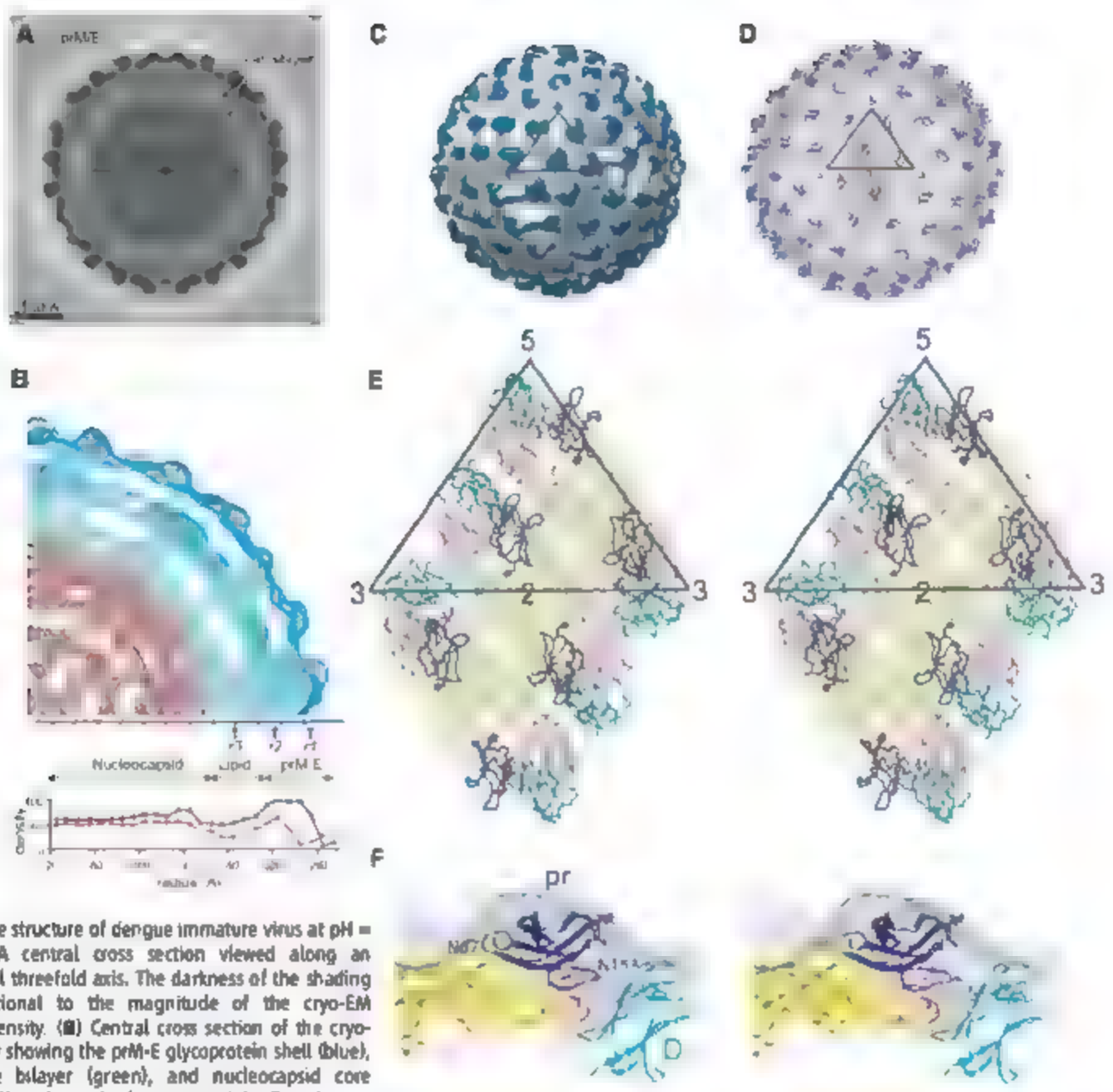


Fig. 2. The structure of dengue immature virus at pH = 6.0. **(A)** A central cross section viewed along an icosahedral threefold axis. The darkness of the shading is proportional to the magnitude of the cryo-EM electron density. **(B)** Central cross section of the cryo-EM density showing the prM-E glycoprotein shell (blue), membrane bilayer (green), and nucleocapsid core (orange). Also shown is the averaged (red) and maximum (blue) density as a function of radius. **(C)** Surface-shaded representation of the low-pH immature virus. The difference density calculated by subtracting the density of the mature virion is shaded in cyan. An icosahedral asymmetric unit is outlined in a gray triangle. **(D)** Pseudo-atomic structure of the virion obtained by fitting the prM-E crystal structure into the cryo-EM density. The E and pr proteins are shown in gray and blue, respectively. The fusion loop is colored in red. The E dimer on the icosahedral twofold axis is shaded in pink, whereas the two monomers of the general-position dimer are shaded in yellow and green. **(E)** Stereodiagram showing the interactions of pr (blue) with E. Domains I, II, and III of the E protein are shown in pink, yellow, and cyan, respectively. Residue Thr^{83} of pr is indicated by a ball, showing its location on the viral surface. The furin cleavage site is located five residues downstream of Thr^{83}. **(F)** Stereoview of the fitted pr and E proteins together with the outline of the density map (gray). Proteins are colored as in (E), and the glycosylation sites on the E molecules are indicated (N67, Asn67, N153, Asn153). The oligosaccharide ligands associated with the crystal structures of prM and E are shown in ball-and-stick format.

lowest energy state that is stable in an acidic environment. In cases in which the cleavage occurs at neutral pH, either on the cell surface or extracellularly, the proteolytic product is trapped in a metastable conformation until it is acidified in the endocytosis pathway and refolds into the most stable form, the postfusion conformation. In other cases in which proteolytic processing occurs in the acidic environment of TGN, there must be a mechanism to prevent the cleaved fusion protein from folding into the postfusion structure. The known method for influenza virus involves a viral proton channel M2, which is inserted into the membranes of the TGN to neutralize the luminal environment (*25*). When the function of M2 is inhibited by amantadine, the HAs expressed on the surface of infected cells are folded into the low-pH-induced postfusion conformation (*26*). Unlike influenza viruses, the flavivirus genome does not encode a protein that might form a transmembrane proton channel. Data presented here suggest a different mechanism for stabilizing the fusion protein during secretion. Although cleavage of prM occurs in the TGN, the proteolytic product pr is retained to prevent fusion. It remains to be tested whether this model can be applied to other viruses, such as alphaviruses, in which the fusion protein E1 is synthesized as a heterodimer with pE2 (*27*) and in which the proteolytic processing of pE2 is also performed by furin in the TGN (*28*, *29*).

The structure of the low-pH immature viruses also provides a basis for drug design to prevent flavivirus infection. Compounds or peptides that bind with high affinity to the virus at the pr-binding site might stabilize the E dimer interface

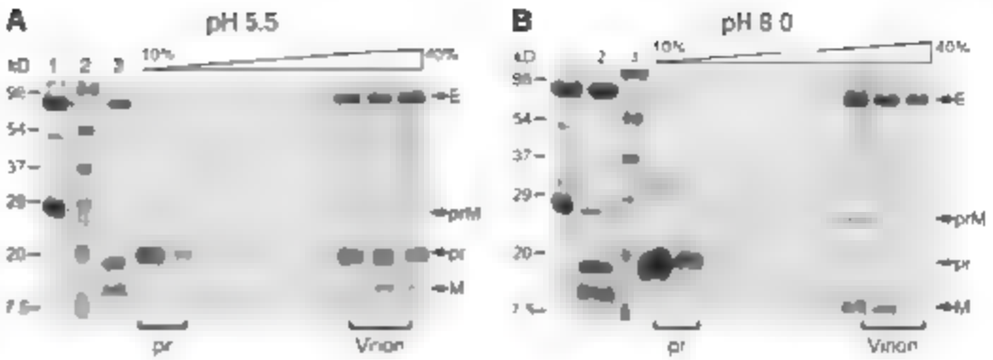


Fig. 3. Association of pr with the virion is pH dependent. Furin-cleaved immature virus was subjected to sucrose sedimentation at (**A**) pH = 5.5 and (**B**) pH = 8.0. Fractions from the gradients were analyzed by SDS-PAGE and silver staining. Soluble pr floats to the top of the gradient, whereas virus-associated pr sediments to the bottom with the virus particles. Lane 1 for (A), uncleaved sample showing the positions of the E and prM proteins; lane 2, molecular weight marker; and lane 3, cleaved sample before loading to the gradients. For (B), lane 1, uncleaved sample; lane 2, cleaved sample before sedimentation; and lane 3, molecular weight marker

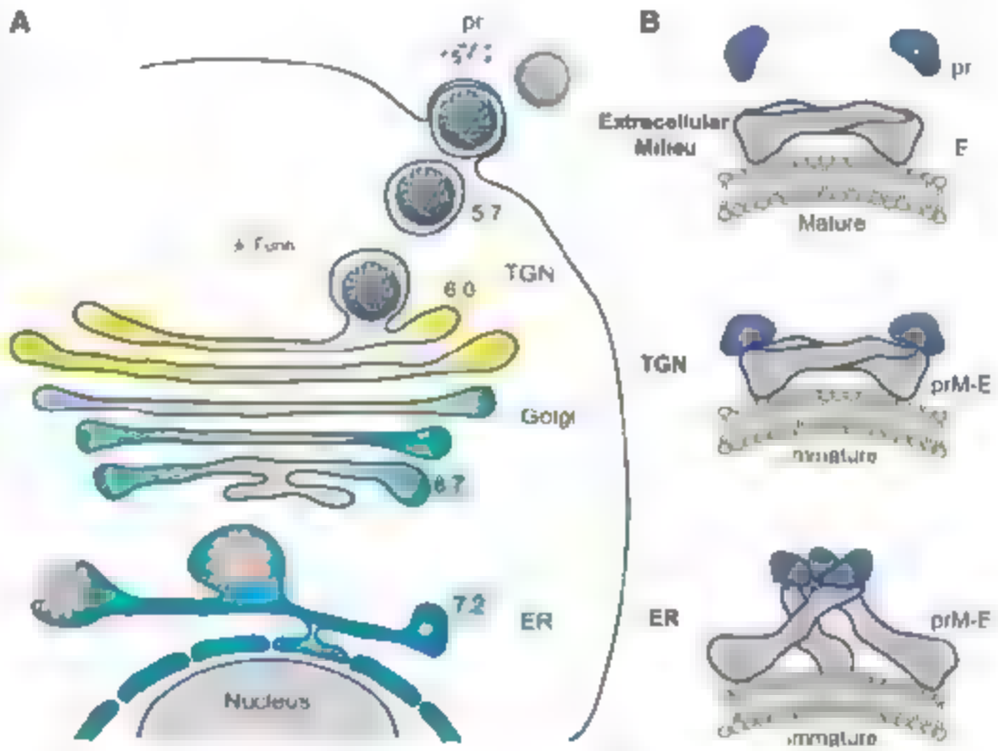


Fig. 4. A model of the flavivirus maturation pathway. (**A**) The conformational changes of the virus particles in the secretory pathway. Immature particles bud into ER as spiky virions and are transported through Golgi into the TGN, where acidification induces a conformational change of the virion. Furin cleavage takes place in the TGN, and pr remains associated until the virion is released to the extracellular milieu. The approximate luminal pH values of the specified cellular compartment are indicated (*30*). (**B**) Configuration of the glycoproteins on the surface of the virion during maturation. The structure of the E protein in the secretory pathway is largely unchanged, except for movements at the hinge between domains I and II. In contrast, the oligomerization states of the glycoproteins are critically dependent on pH. The fusion loops are indicated by red stars.

and block membrane fusion. Because flaviviruses fuse inside the endosome, drugs that target the postfusion conformation require delivery into the low-pH compartment. In comparison, compounds mimicking the pr peptides could act extracellularly and thus would have fewer constraints in the practice of design and application.

References and Notes

1. J. J. Skehel, D. C. Wiley, *Annu. Rev. Biochem.* **69**, 531 (2000).
2. S. Schlesinger, M. J. Schlesinger, in *Fields Virology*, D. M. Knipe, P. M. Howley, Eds. (Lippincott-Raven, Philadelphia, 2001), pp. 895–916.
3. B. D. Lindenbach, C. M. Rice, in *Fields Virology*, D. M. Knipe, P. M. Howley, Eds. (Lippincott-Raven, Philadelphia, 2001), pp. 991–1041.
4. M. Kielian, F. A. Rey, *Nat. Rev. Microbiol.* **4**, 67 (2006).
5. G. Wengler, *J. Virol.* **63**, 2521 (1989).
6. K. Stadler, S. L. Allison, J. Schalich, F. X. Heinz, *J. Virol.* **71**, 8475 (1997).
7. F. A. Rey, F. X. Heinz, C. Mandl, C. Kunz, S. C. Harrison, *Nature* **375**, 291 (1995).
8. R. J. Kuhn *et al.*, *Cell* **108**, 717 (2002).
9. Y. Modis, S. Ogata, D. Clements, S. C. Harrison, *Proc. Natl. Acad. Sci. U.S.A.* **100**, 6986 (2003).
10. Y. Zhang *et al.*, *Structure* **12**, 1607 (2004).
11. R. Kanai *et al.*, *J. Virol.* **80**, 11000 (2006).
12. G. E. Mybakken, C. A. Nelson, B. R. Chen, M. S. Diamond, D. H. Fremont, *J. Virol.* **80**, 11467 (2006).
13. Y. Modis, S. Ogata, D. Clements, S. C. Harrison, *Nature* **427**, 313 (2004).
14. S. Bressanelli *et al.*, *EMBO J.* **23**, 728 (2004).
15. V. B. Randolph, G. Winkler, V. Stollar, *Virology* **174**, 450 (1990).
16. F. Guirakhoo, F. X. Heinz, C. W. Mandl, H. Holzmann, C. Kunz, *J. Gen. Virol.* **72**, 1323 (1991).
17. F. X. Heinz *et al.*, *Virology* **198**, 109 (1994).
18. Y. Zhang *et al.*, *EMBO J.* **22**, 2604 (2003).
19. Y. Zhang, B. Kaufmann, P. R. Chipman, R. J. Kuhn, M. G. Rossmann, *J. Virol.* **81**, 6141 (2007).
20. Materials and methods are available on Science Online.
21. L. Li *et al.*, *Science* **319**, 1830 (2008).
22. S. L. Allison *et al.*, *J. Virol.* **69**, 695 (1995).
23. K. Stiasny, S. L. Allison, A. Marchler-Bauer, C. Kunz, F. X. Heinz, *J. Virol.* **70**, 8142 (1996).
24. S. S. Molloy, L. Thomas, J. K. VanSlyke, P. E. Stenberg, G. Thomas, *EMBO J.* **13**, 18 (1994).
25. F. Ciampor *et al.*, *Virology* **188**, 14 (1992).
26. R. J. Sugrue *et al.*, *EMBO J.* **9**, 3469 (1990).
27. J. M. Wahlberg, W. A. Boere, H. Garoff, *J. Virol.* **63**, 4991 (1989).
28. L. de Curtis, K. Simons, *Proc. Natl. Acad. Sci. U.S.A.* **85**, 8052 (1988).
29. M. Sariola, J. Saraste, E. Kuismanen, *J. Cell Sci.* **108**, 2465 (1995).
30. P. Paroutis, N. Touret, S. Grinstein, *Physiology (Bethesda)* **19**, 207 (2004).

31. We are grateful to S. C. Harrison for helpful discussions, V. C. Livingston for assistance in manuscript preparation; D. C. Marinescu for use of their Linux cluster for image processing; and D. Sedlak, B. Kaufmann, S. Lok, and Y. Zhang for help with virus preparation. The work was supported by an NIH Program Project Grant (AI055672 to R.J.K., M.G.R., and J.C.) and an NIH National Institute of Allergy and Infectious Diseases Region V Great Lakes Center of Excellence for Bio-defense and Emerging Infectious Diseases Research Program award (1-U54-AI-057153 to R.J.K. and M.G.R.). J.C. is a Pew Scholar. The electron density map of the immature virus at pH = 8 has been deposited in the Electron Microscopy Data Bank (accession number EMD-5006). The fitted model of prME molecules has been deposited in the Protein Data Bank (accession number 3C6R).

Supporting Online Material

www.sciencemag.org/cgi/content/full/319/5871/1834/DC1
Materials and Methods
Figs. S1 to S3
Table S1

21 November 2007; accepted 29 February 2008
10.1126/science.1153264

Insect Odorant Receptors Are Molecular Targets of the Insect Repellent DEET

Mathias Ditzen, Maurizio Pellegrino, Leslie B. Vosshall*

DEET (*N,N*-diethyl-meta-toluamide) is the world's most widely used topical insect repellent, with broad effectiveness against most insects. Its mechanism of action and molecular target remain unknown. Here, we show that DEET blocks electrophysiological responses of olfactory sensory neurons to attractive odors in *Anopheles gambiae* and *Drosophila melanogaster*. DEET inhibits behavioral attraction to food odors in *Drosophila*, and this inhibition requires the highly conserved olfactory co-receptor OR83b. DEET inhibits odor-evoked currents mediated by the insect odorant receptor complex, comprising a ligand-binding subunit and OR83b. We conclude that DEET masks host odor by inhibiting subsets of heteromeric insect odorant receptors that require the OR83b co-receptor. The identification of candidate molecular targets for the action of DEET may aid in the design of safer and more effective insect repellents.

Blood-feeding insects transmit many of the world's deadliest diseases. Malaria alone infects an estimated 500 million people annually, leading to the deaths of ~1 million people per year (*1*). In addition to conventional measures of insect control, topically applied insect repellents play a crucial role in protecting humans from blood-feeding insects (*2*). The attraction of mosquitoes to human hosts is largely odor-mediated, with human body emanations such as CO_2, lactic acid, and 1-octen-3-ol acting as strong mosquito attractants (*3*).

Humans have used plant compounds such as eucalyptus and citronella, as well as smoke from incense or burning plant material, for thousands of years to ward off biting insects (*4*). In the 20th century, potent synthetic insect repellents were developed that repelled insects without a strong odor perceptible to humans (*5*, *6*). Among these, DEET is the most commonly used active ingredient of topically applied insect-repellent formulations. It is effective against a wide range of arthropods (*7*, *8*), but its exact mode of action and molecular target are unknown (*8*–*13*). DEET acts as a volatile agent to repel mosquitoes at distances of at least 38 cm from their host (*10*) and repels ticks in vapor phase (*7*). DEET blocks behavioral attraction to lactic acid, a component of human sweat (*11*), and strongly inhibits the electrophysiological activity of lactic acid–sensitive olfactory sensory neurons (OSNs) on the antennae of *Aedes aegypti* (*12*). DEET also appears to have a deterrent effect on feeding (*8*) and exhibits insecticidal properties (*9*). We carried out behavioral and electrophysiological experiments in the malaria mosquito and the fruit fly to elucidate a molecular mechanism of action for the observed olfactory repellency of DEET.

CO_2 and 1-octen-3-ol, emitted in human breath, are potent olfactory attractants for *Anopheles gambiae* in the field (*3*, *14*). Two OSNs housed in capitate peg (cp) sensilla in the mosquito maxillary palp respond with high sensitivity to CO_2 and 1-octen-3-ol (*15*). The large-amplitude spiking cpA cell is tuned to CO_2 and expresses three gustatory receptors (GRs)—GPRGR22, GPRGR23, and GPRGR24—whereas the intermediate-amplitude spiking cpB cell is tuned to 1-octen-3-ol and expresses GPROR8 and the mosquito ortholog of the *Or83b* co-receptor, GPROR7 (*15*).

We performed extracellular electrophysiological recordings on the cp sensilla housing these OSNs to test whether DEET affects peripheral reception of these attractants (Fig. 1). CO_2–evoked responses of the cpA cell were unaffected by DEET (Fig. 1, A and C), suggesting that olfactory transduction mediated by the CO_2 receptor is not affected by the repellent. In contrast, DEET strongly inhibited 1-octen-3-ol–evoked responses mediated by GPROR8 + GPROR7 in the cpB cell (Fig. 1, B and D). The B cell responded to 1-octen-3-ol with a median effective concentration (EC_{50}) of 1.7×10^{-10}, which was shifted to 2.5×10^{-7} in the presence of DEET (Fig. 1D). Because the cpB neuron expresses GPROR7, the mosquito ortholog of the OR83b co-receptor, we carried out a series of experiments in *Drosophila* to determine if DEET acts on *Or83b*-dependent olfactory responses.

Fruit flies avoid a food-baited trap whose entrance is treated with DEET (*16*). To investigate the genetic basis of this repulsion, we adapted a previous two-choice olfactory trap assay (*17*, *18*) (Fig. 2A). In the absence of a food bait and DEET, 72% of the flies entered and distributed equally among the two trap vials, and the remaining flies were found in the starting chamber (Fig. 2B, right two bars). When a filter paper

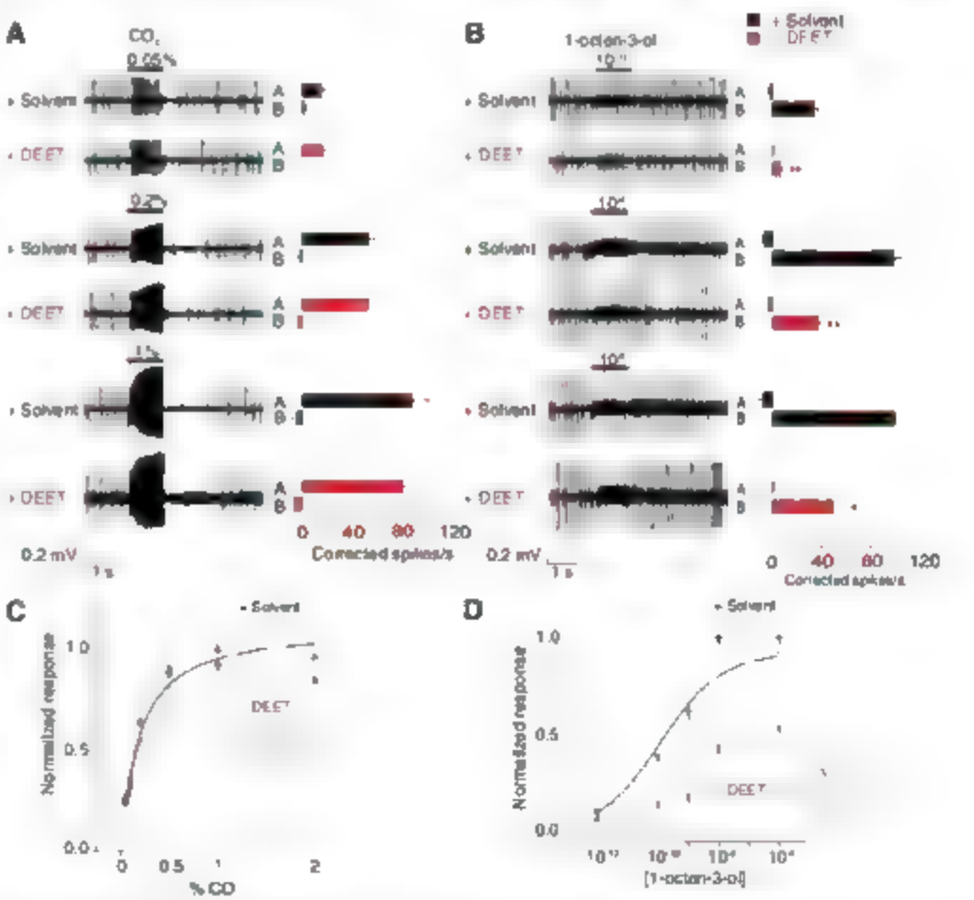


Fig. 1. DEET inhibits mosquito olfactory neuron responses to the attractant 1-octen-3-ol. (**A** and **B**) Recordings from the *A. gambiae* maxillary palp cp sensilla with varying concentrations of CO_2 and 1-octen-3-ol with or without pure DEET. (Left) Representative traces. (Right) Corrected responses of cpA (A) and cpB (B) cells (significance assessed with Mann-Whitney test: $^*P < 0.05$, $^{**}P < 0.01$, unlabeled bars not significantly different; mean ± SEM, $n = 5$ to 7). (**C** and **D**) Dose-response curves of cpA and cpB cell to CO_2 and 1-octen-3-ol with or without DEET (mean ± SEM, $n = 7$ to 13). The chemical structure of DEET is depicted in red.

Laboratory of Neurogenetics and Behavior, The Rockefeller University, 1230 York Avenue, Box 63, New York, NY 10065, USA.

*To whom correspondence should be addressed. E-mail: leslie@mail.rockefeller.edu

lining the entry of one of the two traps was treated with 100% DEET, flies strongly avoided entering this vial (Fig. 2B). If the flies were partially shielded from direct contact with the DEET-treated filter paper by a wire mesh or a perforated polypropylene barrier, avoidance of the DEET-treated trap was reduced (Fig. 2B), and this effect was eliminated when DEET was diluted to 10%, a typical concentration used in insect repellent formulations. Because DEET has been shown to repel airborne mosquitoes (*8*, *10*), we carried out all subsequent behavioral experiments under conditions in which the contact repellent effects of DEET were eliminated by covering 10% DEET with a perforated polypropylene barrier.

To investigate the effect of DEET on food-seeking behavior of *D. melanogaster*, we baited first one and then both traps in the assay with fly food (Fig. 2C). When only one trap contained food (fig. S1A), 85% of flies entered the food vial, 3% entered the empty control vial, and the remaining flies were found in the starting chamber (fig. S1B). When both traps contained food, 93% of flies entered and distributed roughly equally between the two food vials (Fig. 2D, left). However, when the entrance to one baited food vial was treated with 10% DEET, significantly more flies entered the untreated food vial (Fig. 2D, middle), even though DEET was not repellent per se at this concentration (Fig. 2B). When both entrances were treated with 10% DEET, the distribution was again equalized between the two food traps (Fig. 2D, right).

We then asked whether DEET inhibits food attraction by acting peripherally on the olfactory system. Whereas intact flies and those retaining one antenna continued to avoid the DEET-treated food vial (Fig. 2E, left and middle), flies lacking both antennae but retaining secondary olfactory organs, the maxillary palps, entered both food vials, with a slight preference for the DEET-treated side (Fig. 2E, right). We conclude that DEET avoidance requires antennae, which in *Drosophila* house OSNs as well as sensory neurons tuned to humidity, temperature, and mechanical stimuli (*19*, *20*). To investigate whether DEET acts directly on antennal OSNs, we tested the effect of DEET on flies lacking *Or83b*, an essential co-receptor required for the proper trafficking and function of insect odorant receptors in ~80% of antennal OSNs (*17*, *18*). Whereas wild-type flies avoided the DEET-treated food vial, *Or83b*$^{-/-}$ mutants were insensitive to DEET and distributed equally in the two food vials (Fig. 2F). We presume that *Or83b* mutants continue to detect the food bait because CO_2-sensitive neurons and a class of coeloconic OSNs sensing acids and humidity remain functional in *Or83b* mutants (*17*, *18*, *21*) (fig. S2C). Our observation that *Or83b* mutants are DEET-resistant suggests that DEET acts on the olfactory system and requires *Or83b* function for repellency.

To investigate whether DEET blocks all or only a subset of olfactory responses, we recorded food-evoked electrophysiological responses of all *Or83b*-dependent antennal OSNs (Fig. 3A) as well as coeloconic OSNs (fig. S2). Extracellular spiking activity of OSNs in identified basiconic and trichoid sensilla in response to fly food odor was recorded and compared to the same response when food odor was delivered together with DEET. The odor of fly food causes a range of responses in a large number of different OSNs (Fig. 3A, black bars). Although food-evoked responses in most OSNs were not affected by DEET, a subset of neurons showed potentiation (ab1A, ab3B, ab7 ab8) or inhibition (ab1B, ab5, at6) in the presence of DEET (Fig. 3A, red bars). These data in *Drosophila* are in accord with previous experiments in mosquitoes showing complex effects of DEET on odor-evoked responses in OSNs (*14*).

The strongest inhibition of food odor by DEET was observed in the ab5 sensillum, which houses two OSNs: ab5A, which expresses the odorant receptor (OR) *Or82a*, and ab5B, which expresses *Or47a* (*22*). Both ab5A and ab5B OSNs coexpress and require *Or83b* for function (*17*, *18*). To relate the electrophysiological inhibition of ab5 OSNs to behavioral effects of DEET, we examined DEET-mediated inhibition of ab5 OSNs in more detail by using cognate ligands to distinguish ab5A and ab5B. Whereas methyl acetate–induced responses in the ab2 sensillum were not affected by DEET (Fig. 3B, top), both geranyl acetate–induced ab5A responses (Fig. 3B, middle) and 3-methylthio-1-propanol–induced ab5B responses (Fig. 3B, bottom) were strongly inhibited by DEET. The inhibition by DEET of odor-evoked spiking in ab5B was seen across a range of ligand concentrations for both 3-methylthio-1-propanol (Fig. 3C) and pentyl acetate, another ligand of the OR47a + OR83b receptor expressed in ab5B neurons (*22*, *23*) (Fig. 3D).

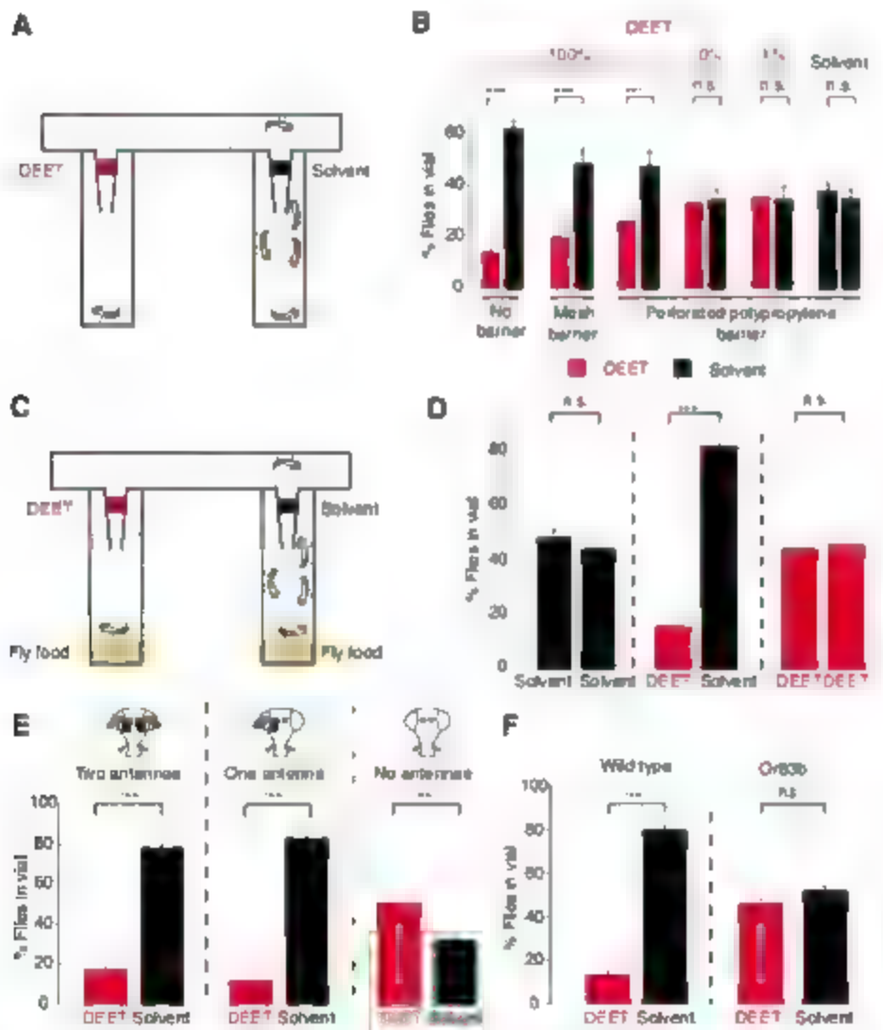


Fig. 2. DEET repellency requires *Drosophila* OR83b. **(A)** Schematic of trap assay without food bait. Entrance to trap is coated with DEET (red) or solvent (black). **(B)** Repellency of varying concentrations of DEET in the trap assay without food bait, with different barriers to impede direct contact with DEET (***$P <$ 0.001, n.s., not significant, Mann-Whitney test; mean ± SEM, n = 11 to 12). **(C)** Schematic of trap assay with fly food bait (yellow). **(D)** Repellency of 10% DEET with perforated polypropylene barrier in the trap assay with food bait (***$P <$ 0.001, n.s., not significant, Mann-Whitney test; mean ± SEM, n = 12, 22, 12). **(E)** Same assay as (D) with surgically de-antennated flies (**$P <$ 0.01, ***$P <$ 0.001, Mann-Whitney test; mean ± SEM, n = 12). **(F)** Same assay as (D) with wild-type and *Or83b*$^{-/-}$ flies (*** $P <$ 0.001, n.s., not significant, Mann Whitney test; mean ± SEM, n = 13, 46).

To investigate whether DEET inhibition of electrophysiological responses of ab5B translated to an effect on behavior, we carried out trap assays with 3-methylthio-1-propanol. Flies showed potent attraction to 3-methylthio-1-propanol (Fig. 3E, left), and this attraction was significantly decreased by DEET (Fig. 3E, right). Odor-evoked activity in the *Or47a*- and *Or83b*-expressing ab5B neuron thus contributes to attractive behavior that can be inhibited by DEET and allows us to correlate a selective electrophysiological effect of DEET with a behavioral phenotype. Consistent with the failure of DEET to inhibit methyl acetate responses of ab2 (Fig. 3B, top), DEET did not alter behavioral responses to methyl acetate in the trap assay (Fig. 3F). We were unable to test the behavioral relevance of ab5A inhibition by DEET because flies were not attracted to geranyl acetate in our trap assays.

We next carried out heterologous expression experiments in which odor-evoked responses of different insect ORs were examined in *Xenopus* oocytes. The functional insect OR is a heteromeric complex of a variable ligand-binding OR subunit and the constant OR83b receptor (*18*, *23*, *24*). Both subunits adopt an atypical topology and share no homology with G protein–coupled receptors (*18*, *25*, *26*). When expressed in heterologous cells, insect ORs mediate odor-evoked increases in intracellular calcium and inward nonselective cation currents (*23*, *24*).

In two-electrode voltage-clamp recordings in oocytes, we confirmed previous results that odor-evoked inward cation currents required coexpression of both OR and OR83b subunits and application of the cognate ligand (*23*) (fig. S2A). Pretreatment of OR47a- and OR83b-expressing oocytes with high concentrations of DEET alone did not generate currents in the oocyte membrane and did not prevent subsequent pentyl acetate–evoked currents in the same oocyte (fig. S2B, upper trace). This suggests that DEET does not have nonspecific effects on biological membranes or membrane proteins. When oocytes stimulated with pentyl acetate were challenged with additional pentyl acetate, no effect was seen on the slow inactivation of the current (fig. S3C). However, when DEET was applied in combination with the cognate ligands of four different ORs, the evoked inward current decreased in a DEET dose-dependent and reversible manner (Fig. 4, A, C, E, and G). DEET inhibited the *Drosophila* OR47a + OR83b receptor expressed in ab5B OSNs (Fig. 4A), two *Anopheles* ORs tuned to human body-odor components (*27*) (Fig. 4, C and E), and the *Anopheles* 1-octen-3-ol receptor, GPROR8 + GPROR7 (*15*) (Fig. 4G). DEET inhibition of the 1-octen-3-ol receptor in heterologous cells is in accord with the in vivo inhibition observed in Fig. 1, B and D. Although DEET inhibited odor-evoked currents in oocytes expressing each of the four insect ORs tested in a dose-dependent manner, the extent of inhibition was dependent on OR + OR83b subunit composition (Fig. 4K).

To determine whether the DEET-induced decrease in odor-evoked current was due to a change in ion permeability, we analyzed current-voltage (*I-V*) relation curves during ligand stimulation in the presence and absence of DEET. The effect of DEET on OR-evoked current was symmetric at positive and negative potentials, and no change in reversal potential was observed (Fig. 4, B, D, F, and H). This suggests a reduction in permeability of the channels affected by DEET, but not a change in ion selectivity.

DEET did not affect currents elicited by activation of the chloride channel CFTR by forskolin (Fig. 4, I and J). Moreover, odor-evoked increases in adenosine 3′,5′-monophosphate (cAMP) mediated through the mouse eugenol receptor (mOR-EG) and detected by activation of CFTR currents (*28*) were also not affected by DEET (Fig. 4L). DEET also inhibited three nonselective cation channels structurally unrelated to insect ORs: mouse TRPM8 (mTRPM8), the heterotrimeric rat olfactory cyclic-nucleotide gated (CNG) channel, and the *Drosophila* ether-a-go-go potassium channel (Fig. 4L). This suggests that DEET may interfere generally with the ionic permeability of a subset of cation channels. We conclude that DEET inhibits odor-evoked currents mediated by selected OR + OR83b complexes.

Electrophysiological and behavioral data presented here suggest that DEET inhibits odor-evoked activation of a subset of insect OR + OR83b complexes, thereby inhibiting the perception of food odors. Because DEET did not uniformly inhibit all *Or83b*-dependent responses in *Drosophila*, it is unlikely that the insect odor-

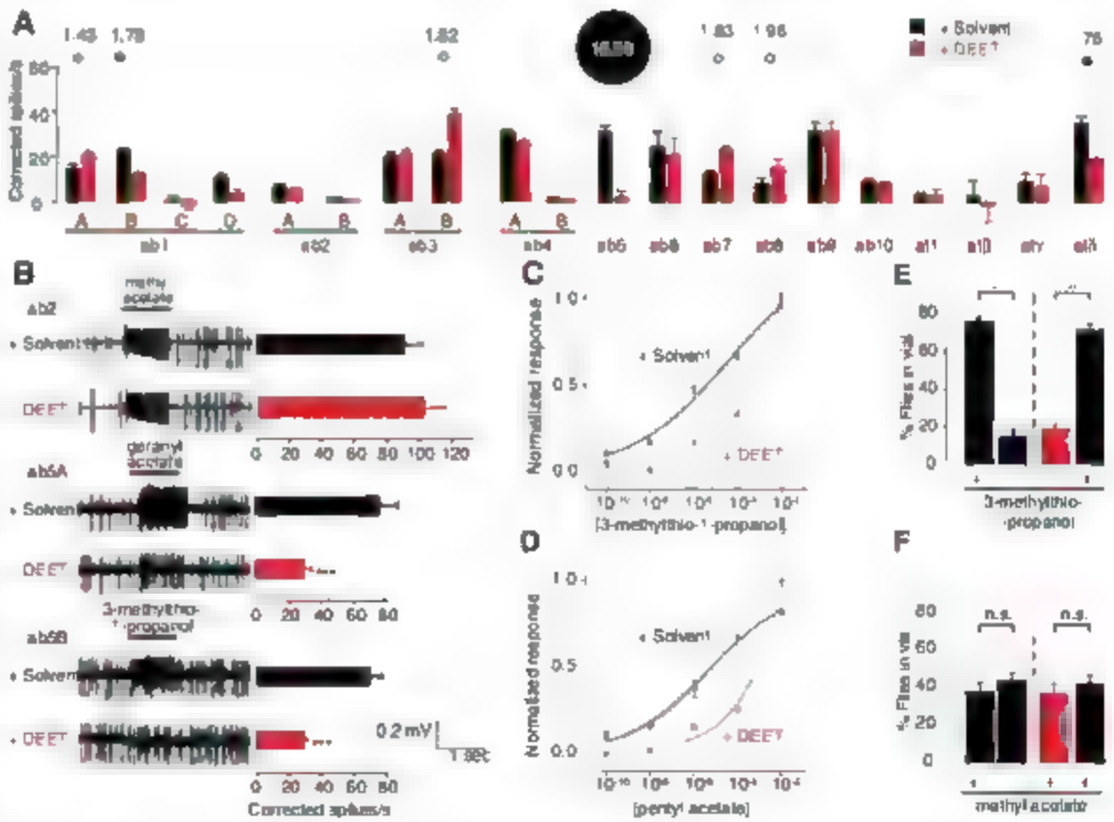


Fig. 3. DEET affects odor-evoked activity of *Drosophila* olfactory sensory neurons. (**A**) Single-sensillum electrophysiology responses of *Or83b*-dependent antennal basiconic (ab) and trichoid (at) sensilla stimulated with food plus solvent (black bars) or food plus DEET (red bars). Data are plotted as mean corrected spikes/s ± SEM (n = 5 to 17 sensilla). Circles above bar graph indicate the fold change in response in the presence of DEET (filled circles, decrease; open circles, increase). (**B**) Representative single-sensillum traces (left) and population responses (right) of ab2 sensilla to methyl acetate at 10^{-5} (top), ab5A neurons to geranyl acetate at 10^{-3} (middle), and ab5B neurons to 3-methylthio-1-propanol at 10^{-5} (bottom) with solvent or DEET (significance assessed with Mann-Whitney test: *** $P < 0.001$, unlabeled bars not significantly different; mean corrected spikes/s ± SEM, n = 8, 8, 10). (**C** and **D**) Dose-response curves of ab5B stimulated with 3-methylthio-1-propanol (C) or pentyl acetate (D), with solvent (black) or DEET (red) (mean ± SEM, n=4). (**E**) (Left) Trap assay in which one vial is baited with pure 3-methylthio-1-propanol (*$P < 0.05$, Mann-Whitney test; mean ± SEM, n = 4). (Right) Repellency of 10% DEET with perforated polypropylene barrier in the trap assay with pure 3-methylthio-1-propanol as bait (***$P < 0.001$, Mann-Whitney test; mean ± SEM, n = 12). (**F**) Same experiment as (E) with pure methyl acetate as bait (n.s., not significant, Mann-Whitney test; mean ± SEM, n = 12, 12).

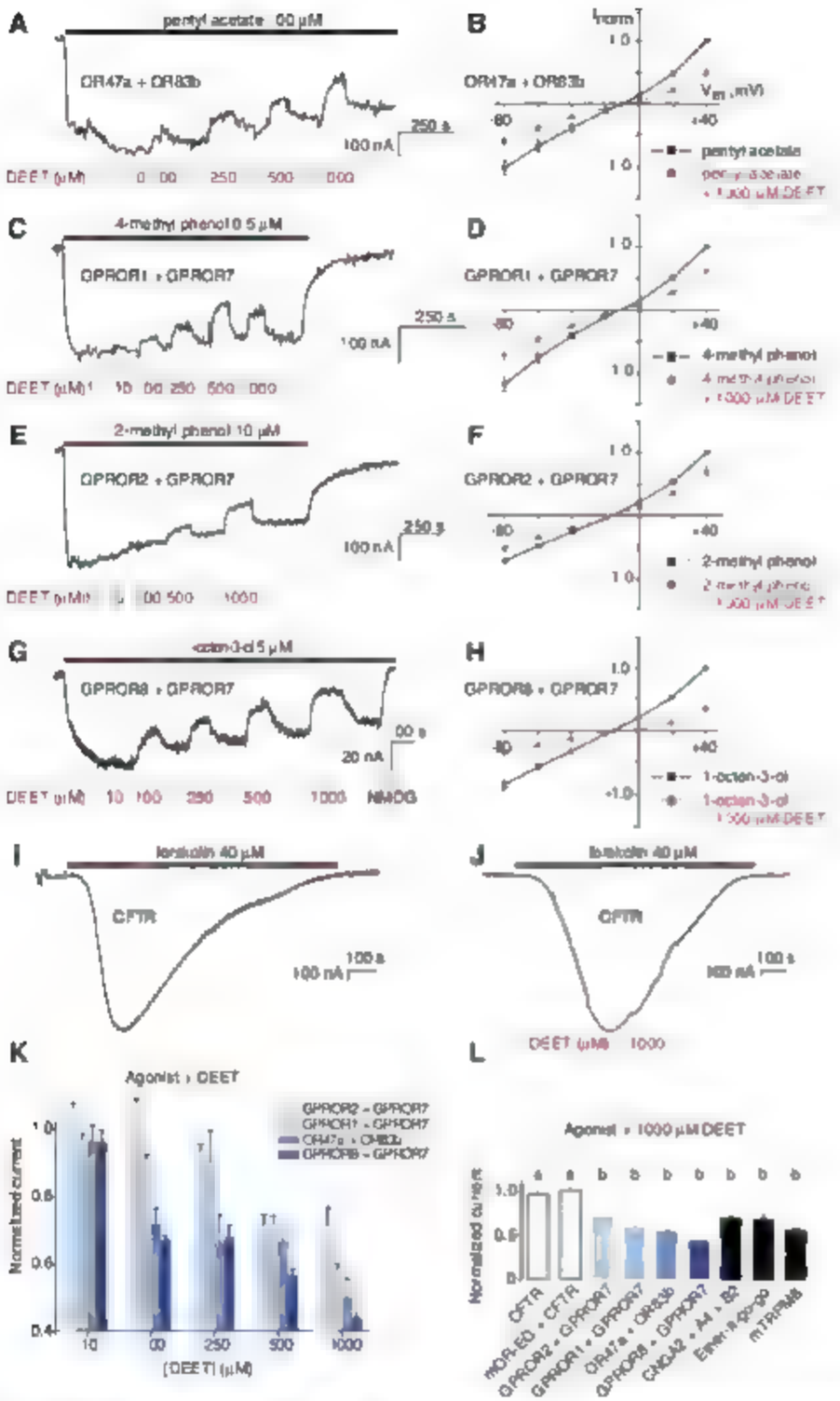


Fig. 4. DEET decreases currents mediated by insect ORs (**A** to **H**) Ligand-evoked currents in the presence of DEET in oocytes expressing OR47a + OR83b (A), GPROR1 + GPROR7 (C), GPROR2 + GPROR7 (E), and GPROR8 + GPROR7 (G). (B), (D), (F), and (H) show current-voltage (*I-V*) curves during ligand stimulation in the absence (black squares) or presence (red circles) of 1000 µM DEET. Current was normalized to the value of +40 mV in the absence of DEET (mean ± SEM, $n = 3$ to 6). (**I** and **J**) Forskolin-evoked currents in the absence (I) or presence (J) of 1000 µM DEET. (**K**) DEET effects on ligand-dependent currents of insect ORs (mean ± SEM, $n = 3$ to 5). Current was normalized to the value of the current in the absence of DEET. (**L**) Normalized stimulus-evoked currents in oocytes expressing various receptors or ion channels in the presence of 1000 µM DEET (CFTR: 40 µM forskolin, mOR-EG + CFTR: 50 µM eugenol; GPROR2 + GPROR7: 10 µM 2-methylphenol; GPROR1 + GPROR7: 0.5 µM 4-methyl phenol, OR47a + OR83b: 100 µM pentyl acetate, GPRORB + GPROR7: 5 µM 1-octen-3-ol; CNGA2+A4+B2: 100 µM cAMP; Ether-a-go-go: voltage steps from −60 mV to +20 mV; mTRPM8: 50 µM menthol). Bars labeled with different letters are significantly different ($P < 0.05$, Kruskal Wallis test with posthoc multiple comparison correction against the CFTR control; mean ± SEM, $n = 4$ to 7).

and co-receptor encoded by OR83b or GPROR7 alone is a direct target of DEET. Although the basis for this selective inhibition by DEET of certain OR + OR83b complexes remains to be investigated, we favor a model in which the subunit composition of the OR + OR83b complex governs both the sensitivity to DEET inhibition and the response properties of the receptor, as previously suggested for *Drosophila* ORs (*22*).

Past attempts to identify novel insect repellent compounds with improved efficacy compared to DEET have been hampered by the absence of a known molecular target for this insect repellent and have relied instead on computational chemistry to explore chemical space around the structure of DEET. We propose that DEET inhibits odor-evoked activation of a subset of OR + OR83b complexes. Accordingly, cell-based assays of behaviorally relevant OR + OR83b complexes could be exploited for high-throughput screening to identify new inhibitory compounds that could represent highly effective insect repellents for use in interrupting infectious disease transmission.

References and Notes

1. J. G. Breman, *Am. J. Trop. Med. Hyg.* **64**, 1 (2001).
2. M. Rowland, T. Freeman, G. Downey, A. Hadi, M. Saeed, *Trop. Med. Int. Health* **9**, 343 (2004).
3. W. Takken, B. G. Knols, *Annu. Rev. Entomol.* **44**, 131 (1999).
4. S. J. Moore, A. Lenglet, in *Traditional Medicinal Plants and Malaria*, M. Wilcox, G. Bodeker, P. Rasoanaivo, Eds. (CRC Press, Taylor and Francis, London, 2004), pp. 343–363.
5. I. H. Gilbert, H. K. Gouck, C. N. Smith, *J. Econ. Entomol.* **48**, 741 (1955).
6. E. T. McCabe, W. F. Barthel, S. I. Gertler, S. A. Hall, *J. Org. Chem.* **19**, 493 (1954).
7. J. F. Carroll, J. A. Klun, M. Debboun, *Med. Vet. Entomol.* **19**, 101 (2005).
8. J. A. Klun, A. Kramer, M. Debboun, *J. Med. Entomol.* **43**, 34 (2006).
9. R. D. Xue, A. Ali, D. R. Barnard, *Exp. Parasitol.* **116**, 201 (2007).
10. C. E. Schreck, I. H. Gilbert, D. E. Weidhaas, K. H. Posey, *J. Econ. Entomol.* **63**, 1576 (1970).
11. E. B. Dogan, J. W. Ayres, P. A. Rossignol, *Med. Vet. Entomol.* **13**, 97 (1999).
12. E. E. Davis, P. G. Sokolove, *J. Comp. Physiol. A* **96**, 223 (1975).
13. S. B. McIver, *J. Med. Entomol.* **18**, 357 (1981).
14. W. Takken, D. L. Kline, *J. Am. Mosq. Control Assoc.* **5**, 311 (1989).
15. F. Lu *et al.*, *Curr. Biol.* **17**, 1533 (2007).
16. M. L. Reeder, P. J. Ganz, J. R. Carlson, C. W. Saunders, *J. Econ. Entomol.* **94**, 1584 (2001).
17. M. C. Larsson *et al.*, *Neuron* **43**, 703 (2004).
18. R. Benton, S. Sachse, S. W. Michnick, L. B. Vosshall, *PLoS Biol.* **4**, e20 (2006).
19. O. Sayeed, S. Benzer, *Proc. Natl. Acad. Sci. U.S.A.* **93**, 6079 (1996).
20. M. J. Kernan, *Pflugers Arch.* **454**, 703 (2007).
21. C. A. Yao, R. Ignell, J. R. Carlson, *J. Neurosci.* **25**, 8359 (2005).
22. E. A. Hallem, M. G. Ho, J. R. Carlson, *Cell* **117**, 965 (2004).
23. T. Nakagawa, T. Sakurai, T. Nishioka, K. Touhara, *Science* **307**, 1638 (2005).
24. E. M. Neuhaus *et al.*, *Nat. Neurosci.* **8**, 15 (2005).
25. M. Wistrand, L. Käll, E. L. Sonnhammer, *Protein Sci.* **15**, 509 (2006).
26. C. Lundin *et al.*, *FEBS Lett.* **581**, 5601 (2007).
27. E. A. Hallem, A. Nicole Fox, L. J. Zwiebel, J. R. Carlson, *Nature* **427**, 212 (2004).

28. S. Kataoka, T. Nakagawa, H. Katayama, K. Touhara, *Biochem. Biophys. Res. Commun.* **305**, 964 (2003)

29. We thank P. Rivkin for technical assistance; K. Lee, J. Klun, K. Touhara, and members of the Vosshall Lab for comments on the manuscript; and P. Howell and M. Q. Benedict of the Centers for Disease Control and Prevention and MR4 for mosquitoes. DNA clones were provided by T.-Y. Chen (ONGs), A. G. Kovacs (CFTR), A. Patapoutian (mTRPM8), G. Wilson (EAG), and L. Zwiebel (GPRORs). M.P. and L.B.V. thank D. Gadsby and Gadsby lab members M. Mense, P. Artigas, M. Reyes, and P. Holl for valuable discussion and advice. Training, and access to instrumentation. M.O. was supported by a Marie-Josee and Henry Kravis Postdoctoral Fellowship. This work was funded in part by a grant to R. Axel and L.B.V. from the Foundation for the NIH through the Grand Challenges in Global Health Initiative and by NIH grant OD008600 to L.B.V. Author contributions: M.O. carried out the experiments in Figs. 1 to 3, M.P. carried out the experiments in Fig. 4, and L.B.V. supervised the work and wrote the paper.

Supporting Online Material
www.sciencemag.org/cgi/content/full/1153121/DC1
Materials and Methods
Figs. S1 to S3
References

19 November 2007; accepted 14 February 2008
Published online 13 March 2008.
10.1126/science.1153121
Include this information when citing this paper.

Aversive Learning Enhances Perceptual and Cortical Discrimination of Indiscriminable Odor Cues

Wen Li,[1]* James D. Howard,[1] Todd B. Parrish,[1,2] Jay A. Gottfried[1,3,4]

Learning to associate sensory cues with threats is critical for minimizing aversive experience. The ecological benefit of associative learning relies on accurate perception of predictive cues, but how aversive learning enhances perceptual acuity of sensory signals, particularly in humans, is unclear. We combined multivariate functional magnetic resonance imaging with olfactory psychophysics to show that initially indistinguishable odor enantiomers (mirror-image molecules) become discriminable after aversive conditioning, paralleling the spatial divergence of ensemble activity patterns in primary olfactory (piriform) cortex. Our findings indicate that aversive learning induces piriform plasticity with corresponding gains in odor enantiomer discrimination, underscoring the capacity of fear conditioning to update perceptual representation of predictive cues, over and above its well-recognized role in the acquisition of conditioned responses. That completely indiscriminable sensations can be transformed into discriminable percepts further accentuates the potency of associative learning to enhance sensory cue perception and support adaptive behavior.

The ability to minimize contact with aversive experience is a hallmark of adaptive behavior. Via mechanisms of associative learning, organisms can use sensory information in the environment to predict impending danger and initiate fight-or-flight responses. The behavioral efficacy of associative learning thus hinges on sensitive and accurate perceptual evaluation of sensory signals. In particular, the ability to discriminate between biologically meaningful cues (e.g., smell of a 175-kg lion) and similar but irrelevant stimuli (e.g., smell of a 3-kg housecat) maximizes an organism's response sensitivity while minimizing hypervigilant and impulsive behaviors.

However, models of associative learning have traditionally focused on delineating the formation of associations between a sensory cue [the conditioned stimulus (CS)] and a biologically salient event [the unconditioned stimulus (US)] (*1*, *2*), paying scant attention to perceptual changes in the CS itself. Several studies have considered how associative learning modifies cue-related tuning profiles in sensory cortex (*3*–*8*), although none has provided concomitant measures of sensory perception. As a consequence, direct links relating learning-induced changes in sensory cortex to perceptual gains in cue discrimination are unavailable, such that the functional importance of these neural effects on behavior remains poorly characterized. To the extent that conditioning can transform indiscriminable sensations into distinct percepts, such a mechanism would constitute a unique and potent means of optimizing adaptive behavior.

We combined functional magnetic resonance imaging (fMRI) with multivariate analytical techniques to explore the impact of aversive olfactory conditioning on perceptual and neural discrimination of predictive odor cues. The use of perceptually identical odor enantiomers (mirror-image molecules differing only in their chiral properties) (*9*, *10*) enabled us to determine whether humans can acquire the ability to distinguish between odorous stimuli that initially smell the same. Twelve healthy human subjects (age range, 22 to 35 years; 8 female) were presented with four enantiomers (two different pairs), one of which (the target CS+, "tgCS+") was repetitively paired with an electric shock (US) during a conditioning phase, whereas its chiral counterpart ("chCS+") was not accompanied by shock (Fig. 1) (*10*). The second pair of odor enantiomers served as nonconditioned control stimuli ("CS−" and "chCS−"). The central prediction was that associative learning would enhance behavioral discrimination of related CS+ odorants, in parallel with reorganization of neural coding in human primary olfactory (piriform) cortex.

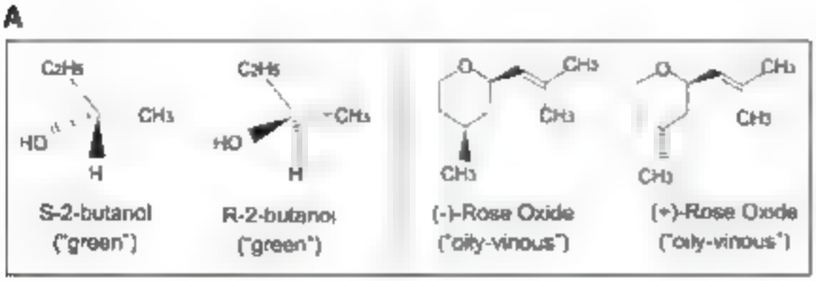
We first examined the behavioral effects of aversive conditioning on perceptual discrimination between the conditioned cue (tgCS+) and its related enantiomer (chCS+). We administered a triangular (triple-forced-choice) odor discrimination test (*10*, *11*) to assess differences in perceived odor identity (i.e., the quality or character of a smell emanating from an odorous object). On each trial, subjects smelled sets of three bottles (two containing one odorant, the third containing its chiral opposite) and selected the odd stimulus. Before conditioning, discrimination accuracy was at chance (33%) for both CS+ and CS− enantiomer pairs, confirming that each pair was initially indistinguishable (Fig. 2A). After conditioning, behavioral accuracy for distinguishing between tgCS+ and chCS+ rose by more than a factor of 2, significantly exceeding both chance and preconditioning performance ($P_s < 0.01$; Wilcoxon test, two-tailed), without any improvement in distinguishing between CS− and chCS−. Subjective ratings of odor intensity, valence, or familiarity (*11*) did not vary across conditions ($P_s > 0.4$), ruling out confounds of the triangular test due to these extraneous variables and accentuating the change in perceived odor identity. Associative learning thus can enhance perceptual discriminability between initially indistinguishable odors, and these effects are specific for the CS+.

We next clarified the neural mechanisms underlying learning-induced perceptual enhancement of the predictive cue (*11*). Because neural representations of odor identity are maintained in posterior piriform cortex (*12*–*14*), and given the highly distributed spatial organization of afferent projections into the piriform region (*15*–*17*), we used multivariate fMRI (*18*, *19*) to test the hypothesis that spatially distributed patterns of neural activity in piriform cortex evoked by tgCS+ and chCS+ would be reorganized as a consequence of associative learning (fig. S1).

By extracting the raw fMRI signal intensity from every activated piriform voxel (fig. S2), we found that the spatial activity patterns in posterior piriform cortex strongly correlated for the CS+ pair (tgCS+:chCS+) and the control pair (CS−:chCS−) before odor-shock learning (Fig. 2B), corresponding to the high perceived similarity within each pair. However, after conditioning, these spatial correlations declined for the CS+ pair, particularly in

[1]Cognitive Neurology and Alzheimer's Disease Center, Feinberg School of Medicine, Northwestern University, Chicago, IL 60611, USA. [2]Department of Radiology, Feinberg School of Medicine, Northwestern University, Chicago, IL 60611, USA. [3]Department of Neurology, Feinberg School of Medicine, Northwestern University, Chicago, IL 60611, USA. [4]Department of Psychology, Northwestern University, Evanston, IL 60208, USA.

*To whom correspondence should be addressed. E-mail: wenli@northwestern.edu.



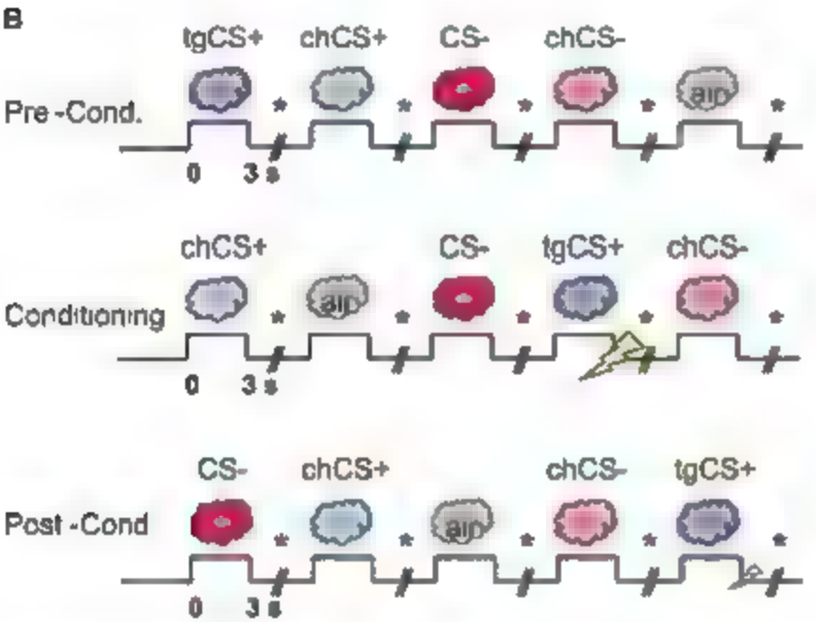


Fig. 1. Experimental paradigm. **(A)** Chemical structures of the enantiomer pairs. **(B)** Learning task. Odorants included a target CS+ (tgCS+) destined for aversive conditioning, its chiral counterpart (chCS+), a nonconditioned control (CS–), and its chiral counterpart (chCS–). A baseline condition consisted of odorless air. Stimuli were delivered during preconditioning, conditioning, and postconditioning sessions. During conditioning, tgCS+ presentation coterminated with electric shock (the US). During postconditioning, the US was presented with tgCS+ on 4 of 19 trials to prevent extinction. On each trial, participants indicated whether odor was present or absent (asterisks). SCR and respiration were continuously recorded.

comparison to the CS pair ($P < 0.05$, Wilcoxon test, two-tailed) (Fig. 2B), in line with the learning-induced behavioral enhancement in odor discrimination between tgCS+ and chCS+. Additional analysis showed that within-pair correlations were significantly higher than across-pair correlations at preconditioning ($P < 0.005$, f 0.66), which suggests that our multivariate technique has satisfactory discriminant validity for distinguishing between odor classes (*11*). Finally the absence of respiratory differences across conditions (*11*) suggests that the imaging effects were not due to sniff-related confounds.

Condition-specific odor maps from one subject (Fig. 3) illustrate how spatial patterns of piriform activity were selectively reorganized from pre- to postlearning for the CS+ pair, whereas the patterns for the CS- pair remained highly coupled. Spatial "difference" maps (Fig. 3, right column) further show that there was minimal signal deviation (light-colored voxels) between CS- and chCS- at both pre- and postconditioning, whereas substantial pattern variation (dark-colored voxels) emerged between tgCS+ and chCS+ postconditioning. Moreover, the response in any given piriform voxel could change in either direction (activation or deactivation); this exemplifies the sensitivity of multivariate pattern-based fMRI approaches to characterizing neural information in human sensory cortex (*20*).

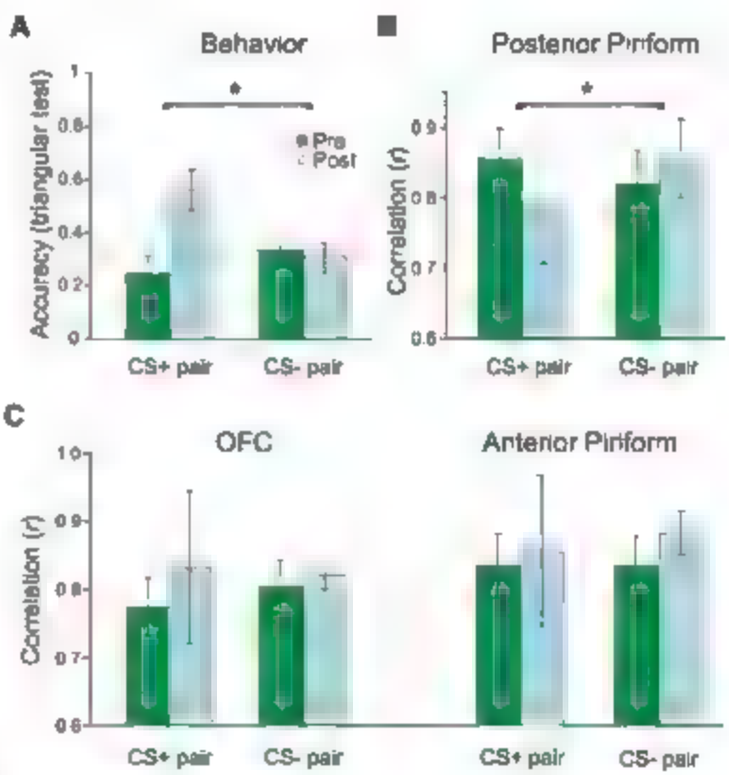


Fig. 2. Parallel enhancement of perceptual and neural discrimination after aversive learning. **(A)** Odor discrimination accuracy was at chance (dashed line) for CS+ and CS– pairs before conditioning, but selectively improved for the CS+ pair after conditioning. Error bars, ±SEM. **(B)** Spatial patterns of fMRI activity in posterior piriform cortex between tgCS+ and chCS+ were highly correlated before conditioning but became more distinct (less correlated), relative to the CS– pair, after conditioning. **(C)** Learning-induced effects on voxelwise spatial activity in OFC (left) and anterior piriform cortex (right) indicated that postconditioning patterns became more correlated (although not significantly) for both CS+ and CS– pairs.

The posttraining changes described above were paralleled by robust evidence for aversive conditioning. First, online physiological measurements (*11*) of the odor-evoked skin conductance response (SCR) revealed significant enhancement to tgCS+ at post- versus preconditioning when compared to the CS odorants (Fig. 4A). These SCR changes, however, were not selective for the

tgCS+, because tgCS+ and chCS+ changes did not significantly differ ($P > 0.2$). In fact, there was a small but nonsignificant SCR increase to chCS+ relative to CS− odorants ($P > 0.4$). Second, learning-induced changes in amygdala and orbitofrontal cortex (OFC) (analyzed using conventional fMRI approaches) (*11*) paralleled the SCR effects. A condition × time interaction (*11*) demonstrated progressive decreases in amygdala activity evoked by tgCS+ (versus CS− odors) as learning proceeded (Fig. 4B and table S1), consistent with prior studies of aversive learning that used visual CS+ stimuli (*21*, *22*). Moreover, comparison of post- to preconditioning revealed increased mean responses to tgCS+ (versus CS− odors) in the OFC bilaterally (Fig. 4, C and D), another region implicated in associative learning (*23*, *24*). Interestingly, fear conditioning partially generalized to the chCS+ odorant, which at reduced threshold ($P < 0.005$ uncorrected) showed similar profiles in amygdala and OFC (table S2). These findings validate the efficacy of our paradigm to induce aversive olfactory learning, thereby supporting the idea that the perceptual and neural changes in sensory discriminability were a consequence of associative learning. The results further show that fear conditioning to an odor cue recruits many of the same regions involved in conditioning to visual and auditory cues, emphasizing the multimodal versatility of these learning networks.

The shock-dependent spatial modifications in posterior piriform cortex were seen in the absence of changes in the magnitude of mean activation. Although the conventional (univariate) fMRI analysis revealed increased mean activity in OFC (Fig. 4C), there was no evidence for similar mean changes in posterior piriform cortex, even at reduced threshold ($P < 0.01$ uncorrected). At the same time, fMRI multivariate (pattern) analysis of OFC (Fig. 2C, left) showed no evidence for enhanced spatial discrimination between the CS+ odorants, but rather suggested further loss of coding specificity (more highly correlated patterns). In fact, the spatial correlation changes for the CS+ pair significantly differed between posterior piriform cortex and OFC ($P = 0.01$, Wilcoxon test), highlighting a regional specificity for the ensemble learning effect. This anatomical/functional double dissociation suggests that fear conditioning recruits functionally distinct networks acting in concert to maximize adaptive behavior: an emotion system (e.g., amygdala and OFC) optimized to detect threat signals with high sensitivity, and a perceptual system (e.g., posterior piriform cortex) optimized to encode signal specificity.

We considered that aversive conditioning could have heightened attention (or arousal) to tgCS+, evoking response changes in olfactory cortex. However, anterior piriform cortex, the purported target of human olfactory attention (*25*), was not modulated in response to associative learning, either in univariate (at $P < 0.01$ uncorrected) or multivariate (Fig. 2C, right) analysis. By comparison, the idea that aversive learning updates odor quality representations in posterior piriform cortex would closely accord with its role in coding odor identity in both animal (*13*, *26*) and human (*12*, *14*) models of olfactory processing. It is therefore unlikely that attention or arousal directly modulates odor coding in posterior piriform, although it remains possible that these mechanisms could mediate olfactory perceptual plasticity indirectly.

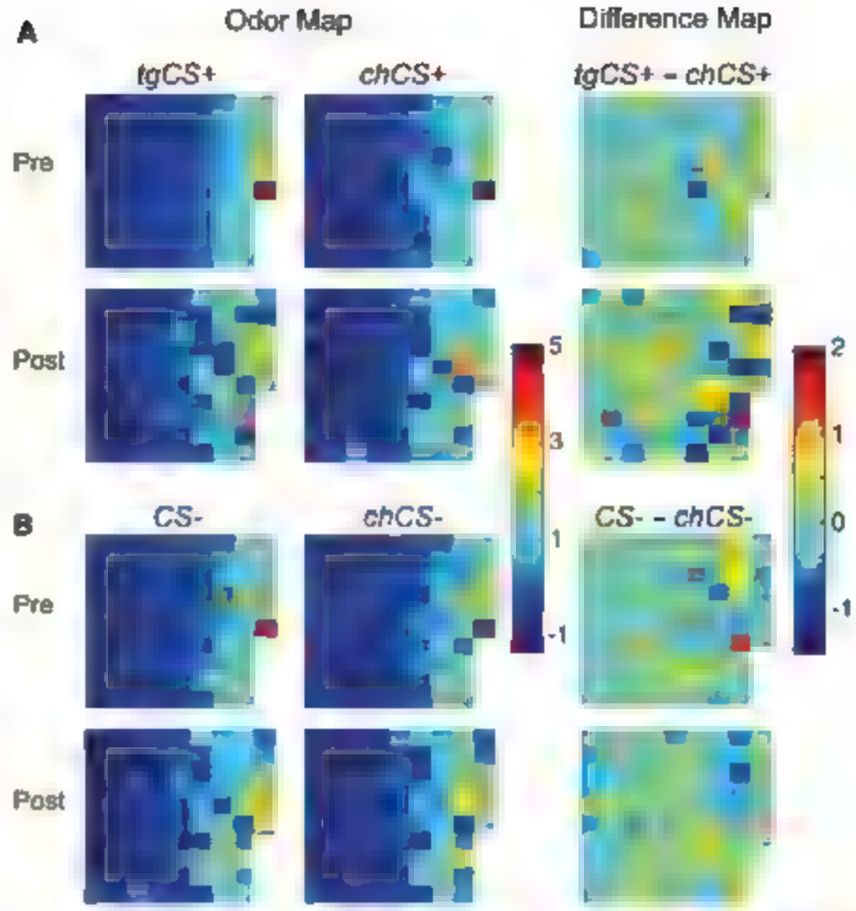


Fig. 3. Spatial maps of posterior piriform activity from one subject. Condition-specific spatial patterns (left two columns) for the CS+ pair (**A**) but not the CS− pair (**B**), diverged after conditioning. Difference maps between odorant pairs (right column) highlight the selective differentiation within the CS+ pair after conditioning. Each square in the grid represents fMRI signal intensity from a different piriform voxel ($n = 86$ voxels), arranged in columns from top left to bottom right, in ascending order of signal intensity for tgCS+ in the preconditioning phase.

Aversive conditioning therefore has a direct influence on how perceptual information about a CS+ is updated in sensory-specific cortex, providing a potent neural substrate to guide the behavioral discrimination of predictive cues. The spatial reorganization of sensory coding in piriform cortex may reflect changes in olfactory receptive-field tuning, leading to improved perception of odor cues, such that unique or "tagged" piriform representations might gain privileged access to critical nodes underlying aversion-minimizing behaviors.

Knowing what to avert presents behavioral challenges that an organism must solve to survive. Prior work on fear conditioning has focused on how the CS comes to produce behavioral (i.e., conditioned) responses, rather than how conditioning alters sensory processing of the CS itself, resulting in perceptual learning and enhanced discrimination. We hypothesize that the substantial effect of emotional experience on perceptual processing in sensory cortices should have a vital impact on adaptive behavior and should thus be considered an indispensable component for models of learning and decision-making (*4*, *27*, *28*). Clinically, our data raise the intriguing possibility that neurobiological derangements in the ability to distinguish between salient cues and perceptually related inconsequential stimuli may underlie the emergence of anxiety disorders characterized by exaggerated sensory sensitivity and hypervigilance. This may provide a unique mechanistic framework for the development of new therapeutic interventions.

References and Notes

1. B. J. Everitt, R. N. Cardinal, J. Hall, J. A. Parkinson, T. W. Robbins, in *The Amygdala: A Functional Analysis*, J. Aggleton, Ed. (Oxford Univ. Press, Oxford, 2000), pp. 353–390.
2. B. W. Balleine, *Physiol. Behav.* **86**, 717 (2005).
3. J. M. Edeline, *Prog. Neurobiol.* **57**, 165 (1999).
4. N. M. Weinberger, *Learn. Mem.* **14**, 1 (2007).
5. D. B. Polley, M. A. Heiser, D. T. Blake, C. E. Schreiner, M. M. Merzenich, *Proc. Natl. Acad. Sci. U.S.A.* **101**, 16351 (2004).
6. F. W. Ohl, H. Scheich, *Curr. Opin. Neurobiol.* **15**, 470 (2005).

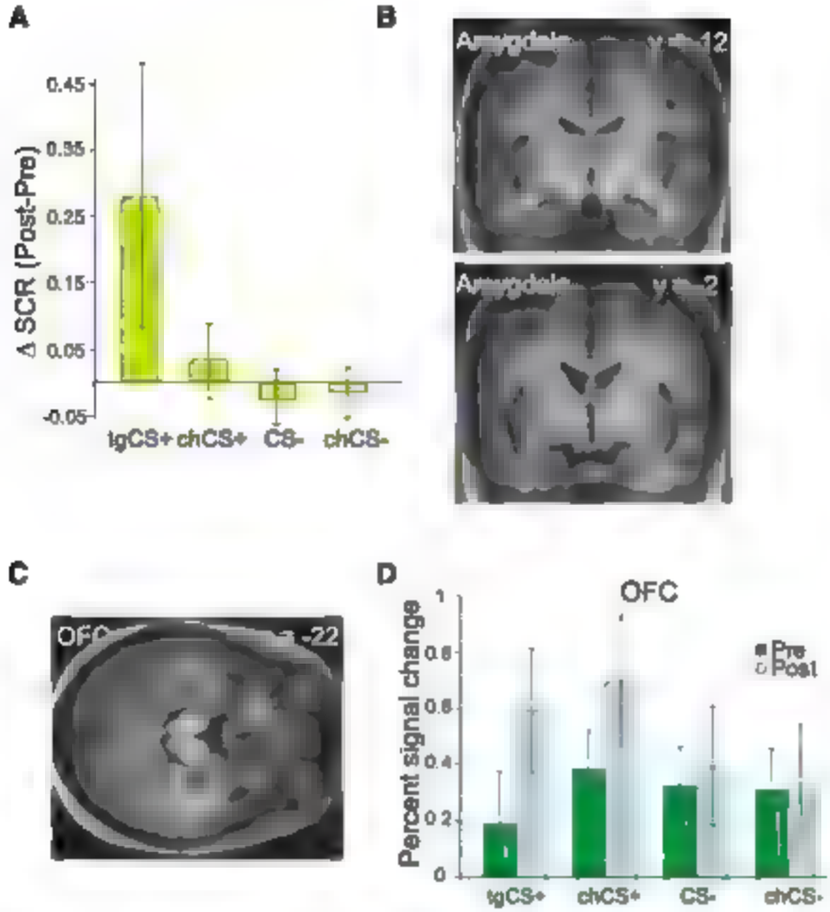


Fig. 4. Effects of aversive olfactory conditioning. **(A)** SCR significantly increased for tgCS+ at post- versus preconditioning relative to CS− stimuli ($P = 0.05$, Wilcoxon test, two-tailed). **(B)** During conditioning, tgCS+–evoked activity in amygdala exhibited significant time-dependent response decline relative to the CS− pair. Activations superimposed on coronal T1-weighted scans (display threshold, $P < 0.001$). **(C)** From pre- to postconditioning, bilateral OFC showed enhanced responses to tgCS+ relative to CS− (axial T1 section; threshold, $P < 0.005$). **(D)** Plots of percent signal change for peak activity in left OFC for each condition.

7 J. S. Morris, K. J. Friston, R. J. Dolan, *Proc. Biol. Sci.* **265**, 649 (1998).
8 E. A. Phelps, J. E. LeDoux, *Neuron* **48**, 175 (2005).
9 C. Linster *et al.*, *J. Neurosci.* **21**, 9837 (2001).
10 M. Laska, P. Teubner, *Chem. Senses* **24**, 161 (1999).
11 See supporting material on Science Online.
12 J. A. Gottfried, J. S. Winston, R. J. Dolan, *Neuron* **49**, 467 (2006).
13 M. Kadohisa, D. A. Wilson, *Proc. Natl. Acad. Sci. U.S.A.* **103**, 15206 (2006).
14 W. Li, E. Lauenberg, T. Parrish, J. A. Gottfried, *Neuron* **52**, 1097 (2006).
15 L. B. Haberly, in *The Synaptic Organization of the Brain*, G. M. Shepherd, Ed. (Oxford Univ. Press, New York, 1998).
16 K. R. Illig, *J. Comp. Neurol.* **488**, 224 (2005).
17 Z. Zou, F. Li, L. B. Buck, *Proc. Natl. Acad. Sci. U.S.A.* **102**, 7724 (2005).
18 J. V. Haxby *et al.*, *Science* **293**, 2425 (2001).
19 S. M. Polyn, V. S. Natu, J. D. Cohen, K. A. Norman, *Science* **310**, 1963 (2005).
20 N. Kriegeskorte, P. Bandettini, *Neuroimage* **38**, 666 (2007).
21 K. S. LaBar, J. C. Gatenby, J. C. Gore, J. E. LeDoux, E. A. Phelps, *Neuron* **20**, 937 (1998).
22 C. Büchel, J. Morris, R. J. Dolan, K. J. Friston, *Neuron* **20**, 947 (1998).
23 G. Schoenbaum, A. A. Chiba, M. Gallagher, *J. Neurosci.* **19**, 1876 (1999).
24 P. C. Holland, M. Gallagher, *Curr. Opin. Neurobiol.* **14**, 148 (2004).
25 C. Zelano *et al.*, *Nat. Neurosci.* **8**, 114 (2005).
26 D. A. Wilson, R. J. Stevenson, *Learning to Smell: Olfactory Perception from Neurobiology to Behavior* (Johns Hopkins Univ. Press, Baltimore, 2006).
27 G. Hall, *Q. J. Exp. Psychol.* **56**, 43 (2003).
28 I. P. L. Mclaren, N. J. Mackintosh, *Anim. Learn. Behav.* **28**, 211 (2000).

29 We thank T. Egner, M. M. Mesulam, J. S. Winston, and R. E. Zinbarg for helpful comments; E. Featherstone, E. Davchev, and V. Djoev for stimulator assembly; and M. Benton for assistance in collecting data. Supported by National Institute on Deafness and Other Communication Disorders grant DC007653 (J.A.G.).

Supporting Online Material

www.sciencemag.org/cgi/content/319/5871/1842/DC1
Materials and Methods
SOM Text
Figs. S1 to S3
Tables S1 and S2
References

9 November 2007; accepted 21 February 2008
10.1126/science.1152837

Electric Fields Due to Synaptic Currents Sharpen Excitatory Transmission

Sergiy Sylantyev,[2*] Leonid P. Savtchenko,[1,2*] Yin-Ping Niu,[3*] Anton I. Ivanov,[4*] Thomas P. Jensen,[1*] Dimitri M. Kullmann,[1] Min-Yi Xiao,[3] Dmitri A. Rusakov[1]†

The synaptic response waveform, which determines signal integration properties in the brain, depends on the spatiotemporal profile of neurotransmitter in the synaptic cleft. Here, we show that electrophoretic interactions between AMPA receptor–mediated excitatory currents and negatively charged glutamate molecules accelerate the clearance of glutamate from the synaptic cleft, speeding up synaptic responses. This phenomenon is reversed upon depolarization and diminished when intracleft electric fields are weakened through a decrease in the AMPA receptor density. In contrast, the kinetics of receptor-mediated currents evoked by direct application of glutamate are voltage-independent, as are synaptic currents mediated by the electrically neutral neurotransmitter GABA. Voltage-dependent temporal tuning of excitatory synaptic responses may thus contribute to signal integration in neural circuits.

Although ion currents through postsynaptic receptors are small (~10^{-11} A), they can exert a lateral voltage gradient (electric field) of 10^4 V/m inside the synaptic cleft (*1*, *2*), which raises the possibility that they can affect the dwell time of electrically charged neurotransmitters (*3*). Does electrodiffusion, therefore, play any role in synaptic transmission?

The excitatory neurotransmitter glutamate is negatively charged at physiological pH (pK 4.4), which implies that postsynaptic depolarization should, in principle, retard its escape from the synaptic cleft (Fig. 1A). AMPA receptor–mediated excitatory postsynaptic currents (AMPAR EPSCs) decay more slowly at positive than at negative holding voltages in dentate basket cells (*4*) and in cerebellar granule cells (*5*). However, this has not been reported for AMPAR EPSCs generated at perisomatic synapses on CA1 or CA3 pyramidal cells (*6*–*8*). We evoked dendritic AMPAR EPSCs in CA1 pyramidal cells by stimulating Schaffer collaterals; the EPSC decay time τ (defined here as the area:peak ratio) increased monotonically with depolarization (Fig. 1B). The ratio between τ recorded at +40 mV and at −70 mV (τ_{+40}/τ_{-70}) was consistently above one [average ± SEM: 2.17 ± 0.09, $n = 49$, $P < 0.001$, (fig. S1A)]. This asymmetry

was independent of the EPSC amplitude, glutamate transport, or recording temperature and could not be accounted for by unknown voltage-dependent properties of receptor antagonists (fig. S1, A and B).

A trivial possible explanation for this phenomenon is that AMPARs themselves have voltage-dependent kinetics. This has indeed been reported for AMPARs activated by brief pulses of glutamate applied to outside-out patches excised from brainstem neurons (*9*) or retinal cells (*10*), but not from hippocampal or dentate granule neurons (*11*). We confirmed that the decay of AMPAR currents evoked by 1 ms/1 mM glutamate pulses in outside-out patches excised from somata ($n = 9$) or dendrites (>150 μm from the soma; $n = 6$) of CA1 pyramidal cells was the same at positive and negative voltages (Fig. 1C). Symmetrical decay kinetics were also observed when the AMPAR density was decreased in the patch with 0.1 μM 6-nitro-7-sulfamoylbenzo[*f*]quinoxaline-2,3-dione (NBQX) (Fig. 1C).

How can the slowed EPSCs at positive voltages observed here be reconciled with the reported voltage independence of the EPSC decay in CA1 pyramidal cells (*6*)? A possible explanation is that EPSCs in previous studies were mainly elicited at proximal synaptic inputs, to ensure optimal voltage clamp (*6*–*8*). Because, in CA1 pyramidal cells, the density of dendritic synaptic AMPARs increases with the distance from the soma (*12*–*14*), the influence of intracleft electric fields on glutamate may be far smaller at perisomatic synapses. To determine whether this explanation is plausible, we simulated the motion of glutamate molecules in the characteristic environment of small hippocampal synapses. Our Monte Carlo approach (*15*) was broadly consistent with previous models (*16*, *17*), with the difference that, at each elemental time step *dt*, each molecule underwent a small displacement because of the local electric field generated by the currents flowing through postsynaptic AMPARs (*18*). Classically, this displacement is given by $Dq(F/RT)\mathbf{E}dt$, where $\mathbf{E}$ is the electric field (*15*), D is the diffusion coefficient (*19*), $q = -1$ for glutamate, F is Faraday's constant, R is the gas constant, and T is temperature. The simulations confirmed that reversal of the AMPAR-mediated synaptic current (by switching to a positive membrane potential) retards the rate of escape of glutamate from the cleft and, consequently, slows the EPSC decay (Fig. 1D and fig. S2, A and B). This effect is consistent with the experimentally observed voltage asymmetry of τ and depends strongly on the number of available synaptic AMPARs (Fig. 1E and fig. S2C). When N is relatively high (>20 open AMPARs at the peak), the effect of electrodiffusion is comparable with that of a twofold change in the glutamate diffusion coefficient (fig. S3). Conversely, the predicted voltage asymmetry of τ is much smaller when N is lower, as expected for proximal synapses.

To test experimentally if reducing the density of activated AMPARs indeed attenuates the voltage asymmetry of τ, we evoked EPSCs at different dendritic locations while visualizing the stimulating pipette position with two-photon excitation microscopy (Fig. 2A). We confirmed that stimulation stochastically evoked synaptic events in dendritic spines (fig. S4). Having documented the τ_{+40}/τ_{-70} ratio in baseline conditions, we blocked a proportion of AMPARs by using 0.1 μM NBQX, which decreased the EPSC amplitude by 40 ± 4% [$n = 7$; (Fig. 2B)]. Because this decrease could also alter voltage clamp conditions of our recordings, we increased the stimulus strength to restore the EPSC amplitude to its baseline value (we thus recruited more local synapses operating at lower AMPAR densities) (Fig. 2B). At all dendritic sites >100 μm from the soma, partial AMPAR blockade substantially reduced the τ_{+40}/τ_{-70} ratio [by 37 ± 5%, $n = 7$, $P < 0.005$; (Fig. 2C)]. In contrast, at more proximal sites, this ratio was mutually much smaller than that at distal sites (1.30 ± 0.06 and 2.30 ± 0.18, $n = 4$ and $n = 7$, respectively, $P < 0.003$), and it was not altered by NBQX application, even though the EPSC amplitude was reduced to the same degree as at distal sites [by 39 ± 3%, $n = 4$, (Fig. 2D)]. Because a reduction in AMPAR density has no effect on receptor kinetics in outside-out patches (Fig. 1C), our observations are consistent with intracleft electric fields decelerating glutamate escape from the cleft when depolarization takes place at distal, but not proximal excitatory synapses in CA1 pyramidal cells.

We designed an alternative approach to test whether the average dwell time of intracleft glu-

[1]Institute of Neurology, University College London, Queen Square, London, WC1N 3BG, UK. [2]Laboratory of Biophysics and Bioelectronics, Dnepropetrovsk National University, Dnepropetrovsk, Ukraine. [3]Institute of Neuroscience and Physiology, Gothenburg University, Gothenburg, Sweden. [4]INMED-INSERM U29, Marseille, France.

*These authors contributed equally to this work.
†To whom correspondence should be addressed. E-mail: d.rusakov@ion.ucl.ac.uk

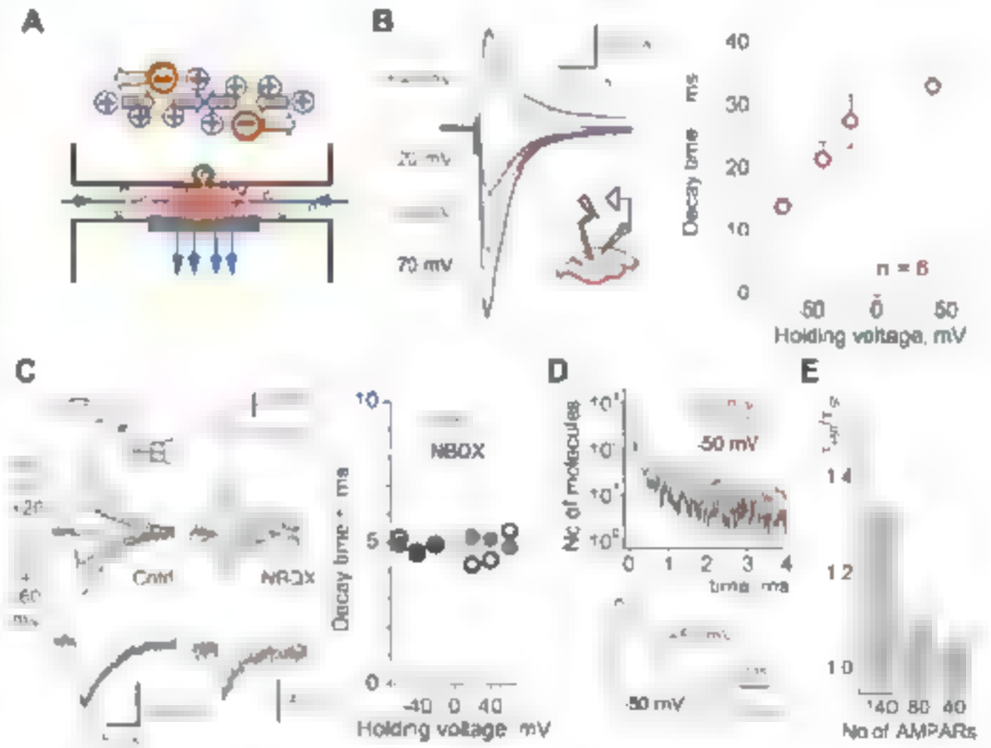


Fig. 1. Interactions between AMPAR currents and glutamate inside the synaptic cleft affect the kinetics of synaptic responses. (**A**) Schematic: negatively charged glutamate molecules experience an electrical force opposite to the cation flux. (**B**) The EPSC decay constant τ (area/amplitude) increases monotonically with depolarization. (left) Example in one cell: AMPAR EPSCs recorded at four holding voltages (color-coded; gray line, response at −70 mV reversed and normalized to that at +40 mV). Plot, voltage-decay relationship in six cells (means ± SEM); dotted line, linear regression. (**C**) (Left) AMPAR currents evoked in outside-out patches by a 1-ms pulse of 1 mM glutamate. (Insets at top) Schematic adapted from (*11*) and the time course of solution exchange; traces, AMPAR responses in baseline conditions (Cntrl, dendritic patch example) and in 0.1 μM NBQX (NBQX, somatic patch example); gray traces, same currents normalized to the amplitude at −60 mV. Graph, average τ ± SEM (Cntrl, filled circles; AMPAR kinetics from $n = 9$ somatic and $n = 6$ dendritic patches were indistinguishable and, therefore, pooled; NBQX, open circles; $n = 7$); dotted lines, global average. (**D**) Monte Carlo simulations predict that postsynaptic depolarization retards escape of glutamate from the cleft (top) and prolongs AMPAR currents (bottom); total 140 AMPARs. (**E**) In simulated AMPAR responses, the ratio τ_{+50}/τ_{-50} (ordinate) depends strongly on the number of synaptic AMPARs (abscissa).

tamate following exocytosis is greater at positive than at negative holding voltages (Fig. 1D). If this is indeed the case, the rapidly dissociating competitive antagonist γ-D-glutamylglycine (γ-DGG) should block a higher fraction of AMPARs at negative voltages than at positive voltages (*20*, *21*), as predicted by modeling (Fig. 3A). First, we confirmed that the kinetics of AMPAR-mediated currents evoked by a brief glutamate

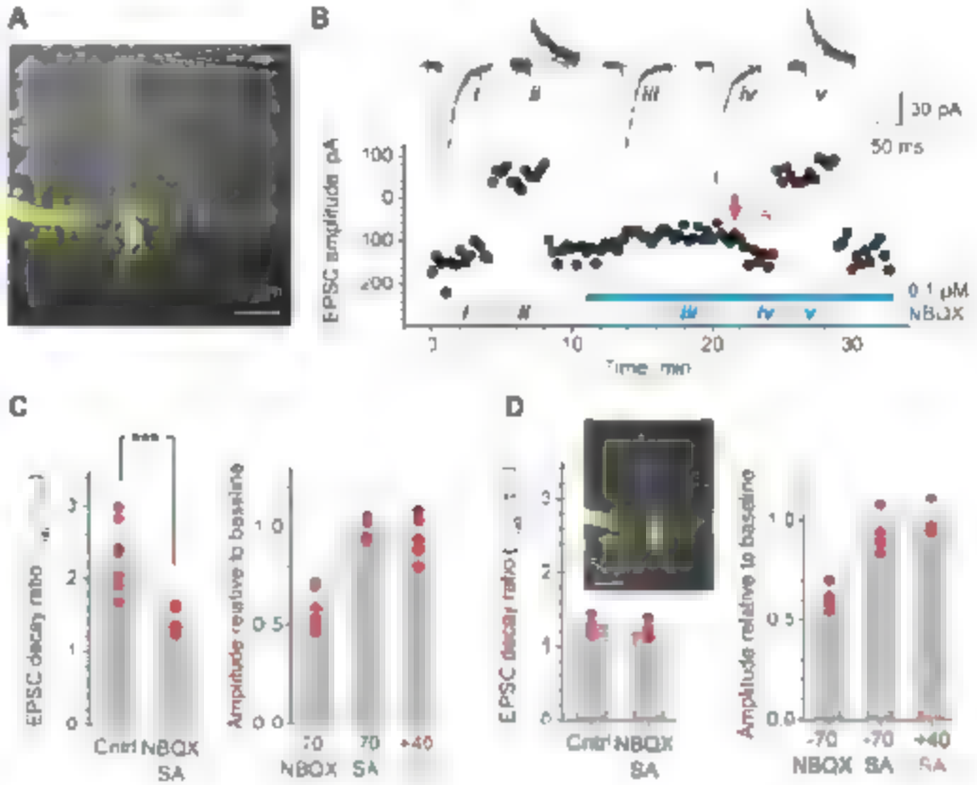


Fig. 2. The voltage-dependent asymmetry of the EPSC decay is prominent at distal, but not proximal, synapses in CA1 pyramidal cells and depends on the AMPAR density. (**A**) Stimulating pipette position (dotted lines) relative to the apical dendrites of a CA1 pyramidal cell (filled with Alexa Fluor 594, $\lambda_x = 800$ nm). (**B**) EPSCs evoked at a remote dendritic site in one cell, as shown in (A), and recorded at −70 mV and +40 mV, before and after application of 0.1 μM NBQX (blue segment); dots, EPSC amplitude; gray segments correspond to the average traces, as indicated by roman numerals; red arrow, stimulus amplitude adjustment (SA). See text for details. (**C**) Summary of experiments shown in (B). The EPSC decay ratio τ_{+40}/τ_{-70} is reduced by 0.1 μM NBQX from 2.30 ± 0.18 to 1.44 ± 0.07 (left) ($n = 7$, $P < 0.005$), whereas the EPSC amplitude is 0.99 ± 0.02 and 0.96 ± 0.04 of its baseline values at −70 mV and +40 mV, respectively. (Right) Bars show average amplitude relative to baseline at the indicated holding voltage. (**D**) In proximal synapses (inset), application of NBQX has no effect on the τ_{+40}/τ_{-70} ratio (baseline, 1.30 ± 0.06; NBQX, 1.22 ± 0.06, $n = 4$; n.s.); notation is as in (C).

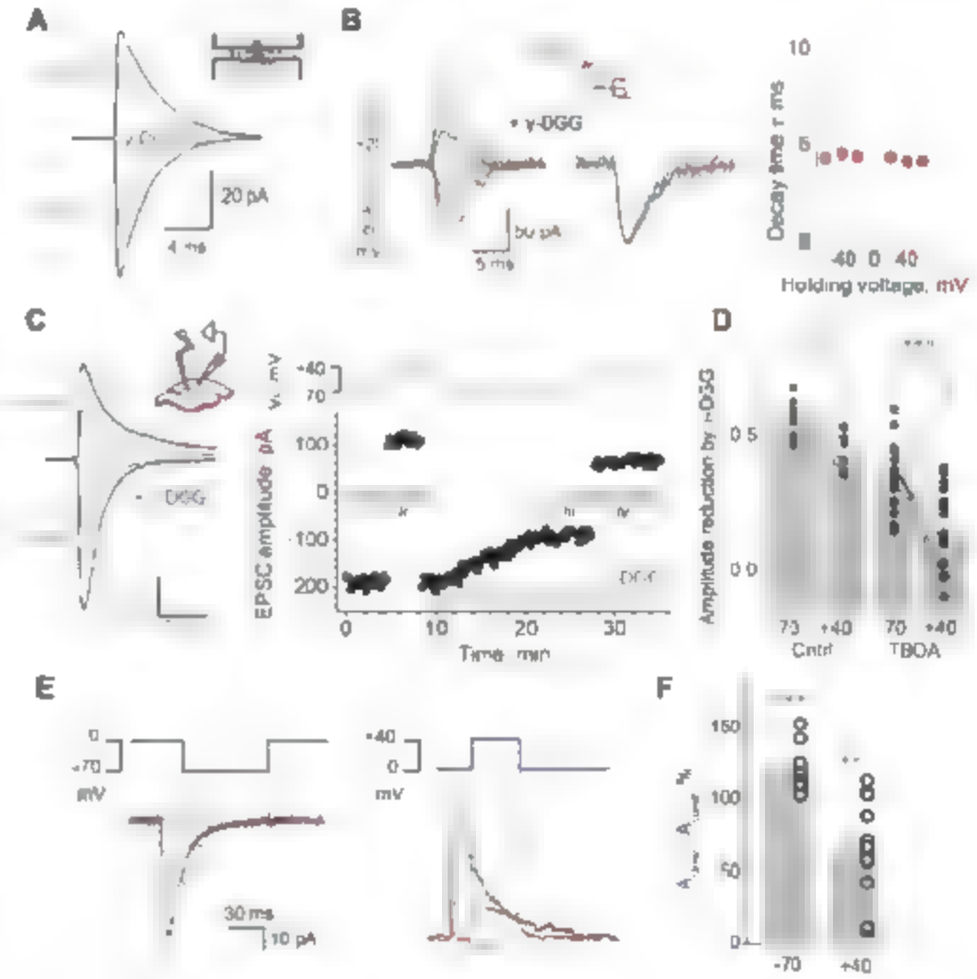


Fig. 3. Postsynaptic membrane voltage affects glutamate escape from the cleft. (**A**) Monte Carlo simulations of AMPAR EPSCs [kinetics in accordance with (*28*), 2000 glutamate molecules released] predict that 0.5 mM γ-DGG should be less effective at positive than at negative holding voltages. (**B**) Outside-out patch experiments showing that interactions between AMPARs and γ-DGG are voltage-independent; notation is as in Fig. 1C. (**C**) In slices, reduction of AMPAR EPSCs by 0.5 mM γ-DGG is smaller at +40 mV than at −70 mV; traces, example from one cell (roman numerals correspond to gray segments in the plot). Plot, the corresponding amplitude time course; (top) holding voltage; blue segment, application of 0.5 mM γ-DGG; gray segments, trace averaging. (**D**) Summary of experiments shown in (C); bars, average values; connected dots, data points from individual cells. Average amplitude reduction by γ-DGG: 51 ± 3% at −70 mV and 43 ± 3% at +40 mV ($n = 8$, $P < 0.01$); in the presence of 50 μM TBOA: 35 ± 3% and 16 ± 4%, respectively ($n = 18$, $P < 0.005$). (**E**) Switching the holding voltage rapidly from the AMPAR reversal potential (~0 mV) unmasks the voltage dependence of response kinetics; one-cell examples. Black, EPSC at −70 mV (left) or +40 mV (right); red trace, synaptic response during the holding-voltage jump (blue) minus the response to the voltage jump alone (no stimulus). (**F**) Summary of experiments shown in (E). Ratio between the average current amplitude over a 10-ms interval starting at 5 to 10 ms after the jump onset (A_{jump}; gray segments in (E)) and the current value over the same time window when the cell is held at either −70 mV or +40 mV throughout the sweep (A_{const}); abscissa, holding voltage in mV.

pulse in outside-out patches in 0.5 mM γ-DGG were voltage-independent (Fig. 3B). In keeping with our prediction, the reduction of AMPAR EPSC amplitudes by 0.5 mM γ-DGG was 20 ± 7% greater at negative, than at positive, holding voltages [$n = 8$, $P < 0.02$, (Fig. 3, C and D)]. Blocking glutamate uptake with 50 μM threo-β-benzyloxyaspartate (TBOA), to rule out any possible contribution of voltage-dependent transporters, increased this difference to 45 ± 16% [$n = 18$, $P < 0.001$, (Fig. 3D)] while reducing the overall effects of 0.5 mM γ-DGG, probably because of an increase in the ambient level of glutamate. In contrast, 0.1 μM NBQX reduced the EPSC amplitude equally at both positive and negative holding voltages, in accordance with modeling predictions (fig. S5). These results further confirm that the sign of the synaptic current influences the rate of escape of glutamate from the synaptic cleft.

Finally, if synaptic currents do influence glutamate diffusion, holding the postsynaptic cell at the receptor reversal potential (zero current) during and immediately after presynaptic glutamate release should abolish the effect of voltage on EPSC decay. We therefore switched the holding voltage from the AMPAR reversal potential (0 mV) to either −70 mV or +40 mV during the EPSC decay phase (Fig. 3E) and compared the outcome with the EPSCs recorded without the voltage jump. The voltage-jump responses showed a 20 ± 5% slower decay at −70 mV ($P < 0.005$, $n = 11$), and a 34 ± 10% faster decay at +40 mV [$P < 0.01$, $n = 12$; (Fig. 3F)] (*18*). The EPSC waveform was thus influenced by the recent history of current flow (*22*), consistent with electrodiffusion.

Does electrodiffusion of glutamate influence the activation of other receptors? High-affinity NMDA receptors (NMDARs) both within and outside synapses may be activated by synaptic releases of glutamate (*23*, *24*). This, in addition to voltage-dependent blockade by Mg^{2+}, is likely to mask the effects of electrodiffusion on NMDAR responses. Nevertheless, we observed a voltage-dependent asymmetry of NMDAR EPSCs in the absence of extracellular Mg^{2+} in cultured hippocampal neurons, which are not surrounded by dense neuropil (Fig. 4A and figs. S6 and S7). This was again consistent with modeling (fig. S8). In contrast, a relatively low concentration of glutamate applied diffusely to the dendrites—to activate NMDARs at a lower density—evoked responses that showed voltage-independent kinetics, but otherwise were similar to those evoked synaptically (Fig. 4A, and fig. S9). Furthermore, reduction of NMDAR EPSCs by the fast-dissociating competitive antagonist D-aminoadipate (D-AA) was greater at negative than at positive voltages (fig. S10). Finally, both slowly and fast dissociating NMDAR antagonists, 0.4 μM 3-(2-carboxypiperazin-4-yl)propyl-1-phosphonic acid (D-CPP) and 50 μM D-AA, reduced the voltage-dependent asymmetry of τ (fig. S11).

These phenomena should play no role in the activation of γ-aminobutyric acid type A ($GABA_A$) receptors because GABA is a zwitterion. We tested this prediction in neuronal cultures, again, to avoid the confounding effects of extrasynaptic and/or tonically active $GABA_A$ receptors in slices. Although $GABA_A$ receptor–mediated inhibitory postsynaptic currents (IPSCs) did decelerate at positive voltages (with respect to the Cl^- reversal potential), the responses were symmetrical when GABA transporters were blocked with 25 μM SKF-89976A (Fig. 4B). Thus, we observed no evidence that electric fields affect the synaptic dwell time of GABA.

Electrodiffusion of glutamate may therefore explain, at least in part, why AMPAR EPSCs at some central synapses are retarded by depolarization (*4*, *5*) and why EPSCs recorded locally at distal dendrites of CA1 pyramidal cells have faster decays than those at proximal dendrites (*12*). The extent of this phenomenon is likely to vary among synapses, depending, for instance, on the density or numbers of synaptic receptors. Although electrodiffusion is thus a fundamental feature of AMPAR-mediated synaptic transmission, does it play a physiologically significant role in synaptic signal integration? Distal dendrites of pyramidal neurons can undergo extensive depolarization (including local spiking) without exciting the soma (*25*) (Fig. 4C), and even modest changes in τ because of electrodiffusion, in principle, should affect the time interval over which near-coincident inputs trigger an action potential. Indeed, simulations with a NEURON (*26*) model of a CA1 pyramidal cell (*12*, *27*) suggest that, for an arbitrary sample of Schaffer collateral input locations (Fig. 4C), a ~20% shortening of the synaptic conductance decay could reduce the coincidence detection window by 52 ± 6% [$n = 16$, $P < 0.001$; (Fig. 4,

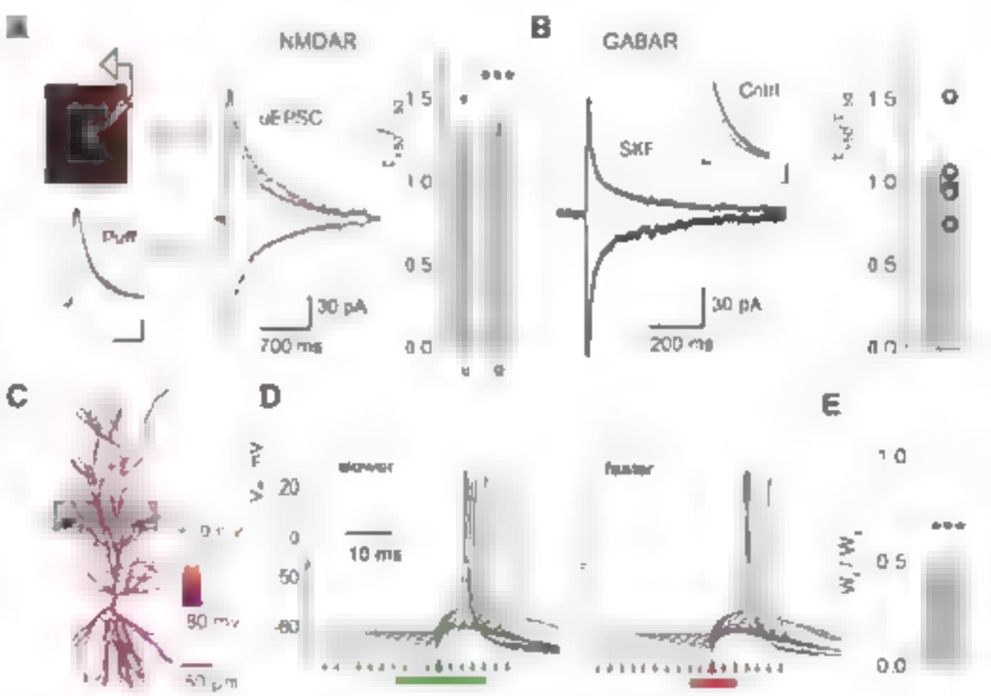


Fig. 4. Effects of electrodiffusion depend on the neurotransmitter charge and may affect signal integration properties in hippocampal neurons. **(A)** NMDAR-mediated EPSCs recorded in culture at zero Mg^{2+} (unitary EPSC, uEPSC); traces from one cell; magenta, response at −50 mV normalized to that at +50 mV. Puff, responses to a pressure pulse of glutamate (scale values apply throughout). Graph, average τ_{+50}/τ_{-50} ratios (± SEM) for unitary (u) and evoked (e) NMDAR EPSCs: 1.29 ± 0.07 and 1.37 ± 0.08 ($n = 6$, $P < 0.02$ and $n = 15$, $P < 0.001$, respectively); see also figs. S6 to S11. **(B)** Synaptically evoked $GABA_A$ receptor–mediated responses in culture show no voltage-dependent asymmetry in the presence of the GABA uptake blocker SKF-89976A (25 μM); Cntrl, responses before GABA uptake blockade; inward and outward current correspond to, respectively, 30 mV below and above the Cl^- reversal potential corrected for junction potential (*18*), other notation is as in (A). Graph, summary, average τ_{+50}/τ_{-50}: 1.05 ± 0.10 ($n = 6$). **(C)** Diagram, a detailed NEURON model of a CA1 pyramidal cell (*12*, *27*); arrows, an example of two synapses that could strongly depolarize local dendrites (producing local spikes) without exciting the soma (V_m color-coded); gray shadow, dendritic area of synaptic pairs tested in (D). **(D)** Simulations illustrating coincidence detection for two synaptic inputs: one activated at time zero (black arrow) and the other activated at 3-ms intervals before and after (gray arrows). During the coincidence detection time window (green and red segments), summation generates a spike. (left and right) The synaptic conductance decay set at 10 ms (slower) and 8 ms (faster), respectively, to reflect the effect of electrodiffusion. **(E)** Summary of experiments shown in (D) for arbitrarily selected synaptic input pairs (located within the gray area in (C)); the coincidence detection window for faster versus slower inputs (W_f/W_s); mean ± SEM: 0.48 ± 0.06 ($P < 0.001$, $n = 38$).

D and E)]. Another potential consequence of electrodiffusion is that local postsynaptic depolarization, by extending the dwell time of intracleft glutamate (Fig. 1D), may enhance activation of NMDARs. This is likely to interact synergistically with the depolarization-dependent attenuation of postsynaptic glutamate transport and relief of Mg^{2+} block, thus potentially facilitating induction of NMDAR-dependent synaptic plasticity.

References and Notes

1. J. C. Eccles, J. C. Jaeger, *Proc. R. Soc. Lond. B. Biol. Sci.* **148**, 38 (1958).
2. L. P. Savtchenko, S. N. Antropov, S. M. Korogod, *Biophys. J.* **78**, 1319 (2000).
3. A. Takeuchi, N. Takeuchi, *J. Neurophysiol.* **22**, 395 (1959).
4. T. B. Knowles, R. Dingledine, *Hippocampus* **5**, 151 (1995).
5. L. Cathala, N. B. Holderith, Z. Nusser, D. A. DiGregorio, S. G. Cull-Candy, *Nat. Neurosci.* **8**, 1310 (2005).
6. S. Hestrin, R. A. Nicoll, D. J. Perkel, P. Sah, *J. Physiol.* **422**, 203 (1990).
7. C. McBain, R. Dingledine, *J. Neurophysiol.* **68**, 16 (1992).
8. P. Jonas, G. Major, B. Sakmann, *J. Physiol.* **472**, 615 (1993).
9. I. M. Raman, L. O. Trussell, *Biophys. J.* **69**, 1868 (1995).
10. M. L. Vituki, S. H. Morlwe, E. Harvet, *J. Physiol.* **549**, [illegible].
11. D. Colquhoun, P. Jonas, B. Sakmann, *J. Physiol.* **458**, 261 (1992).
12. J. C. Magee, E. P. Cook, *Nat. Neurosci.* **3**, 895 (2000).
13. B. K. Andrasfalvy, J. C. Magee, *J. Neurosci.* **21**, 9151 (2001).
14. D. A. Nicholson *et al.*, *Neuron* **50**, 431 (2006).
15. L. P. Savtchenko, D. A. Rusakov, *Proc. Natl. Acad. Sci. U.S.A.* **104**, 1823 (2007).
16. S. Raghavachari, J. E. Lisman, *J. Neurophysiol.* **92**, 2456 (2004).
17. K. M. Franks, T. M. Bartol Jr., T. J. Sejnowski, *Biophys. J.* **83**, 2333 (2002).
18. Materials and methods are available as supporting material on Science Online.
19. T. A. Nielsen, D. A. DiGregorio, R. A. Silver, *Neuron* **42**, 757 (2004).
20. J. D. Clements, R. A. J. Lester, G. Tong, C. E. Jahr, G. L. Westbrook, *Science* **258**, 1498 (1992).
21. J. M. Christie, C. E. Jahr, *J. Neurosci.* **26**, 210 (2006).
22. K. L. Magleby, C. F. Stevens, *J. Physiol.* **223**, 151 (1972).
23. M. Arnth-Jensen, D. Jabaudon, M. Scanziani, *Nat. Neurosci.* **5**, 325 (2002).
24. A. Scimemi, A. Fine, D. M. Kullmann, D. A. Rusakov, *J. Neurosci.* **24**, 4767 (2004).
25. M. Hausser, N. Spruston, G. J. Stuart, *Science* **290**, 739 [illegible].
26. M. L. Hines, N. T. Carnevale, *Neuroscientist* **7**, 123 (2001).
27. M. Migliore, *J. Comput. Neurosci.* **14**, 185 (2003).
28. J. L. Wadiche, C. E. Jahr, *Neuron* **32**, 301 (2001).
29. We thank B. Gustafsson and E. Hanse for their comments and support. This work was supported by the Wellcome Trust, the Medical Research Council (UK), European Union (Prommemoria S12012) and Human Frontier Science Program (RGP50/2006), and also by the Swedish Research Council, the Swedish Society of Medicine and the Göteborg Medical Society.

Supporting Online Material

www.sciencemag.org/cgi/content/full/319/5871/1845/DC1
Materials and Methods
Figs. S1 to S11
References

18 December 2007; accepted 22 February 2008
10.1126/science.1154330

Rule Learning by Rats

Robin A. Murphy,[1]* Esther Mondragón,[2]* Victoria A. Murphy[2]

Using rules extracted from experience to solve problems in novel situations involves cognitions such as analogical reasoning and language learning and is considered a keystone of humans' unique abilities. Nonprimates, it has been argued, lack such rule transfer. We report that *Rattus norvegicus* can learn simple rules and apply them to new situations. Rats learned that sequences of stimuli consistent with a rule (such as XYX) were different from other sequences (such as XXY or YXX). When novel stimuli were used to construct sequences that did or did not obey the previously learned rule, rats transferred their learning. Therefore, rats, like humans, can transfer structural knowledge from sequential experiences.

The ability to extract generalizable rules from specific experiences is a fundamental attribute of human higher cognitive functioning (*1*, *2*). For instance, human language learning relies on learning grammatical rules that allow the English speaker to discriminate "the dog bit the woman" (subject-verb-object) from "bit the woman the dog" (verb-object-subject). Rules can also be transferred to newly encountered items, contributing to the understanding and production of new sentences. Cross-linguistic differences in grammatical rules serve to highlight the challenge facing the developing child: that phrase structure rules must be learned. Language is only one of many cognitive domains where problems of this type are present, raising the issue of the nature of the underlying cognitive substrates supporting rule learning in general.

Rule learning has been investigated in different species and stages of human development (*3*, *4*). Pre-linguistic infants (*5*), primates (*6*), and even some birds (*7*, *8*) can learn rule-like temporal structures. Starlings, for example, are able to learn sequential structures made up of segments of birdsong and can recognize whether patterns are consistent with those in a training set. Although the question of whether the birds use recursion to solve the discrimination is controversial, they do appear to use a form of rule (*8*, *9*).

We studied the ability of rats to learn and transfer rules. We developed a procedure to test rule-transfer learning in the rat, an animal that, it has been argued, might have limited or perhaps even no ability parallel to that used by humans (*3*, *5*). Rats were trained with three-element sequences (such as ABA or BAB) paired with food. We asked whether they would learn something about the overall pattern of cues (such as XYX). At least three cognitive abilities are required in order to learn this rule and apply it in novel situations. The first is an ability to learn that a sequence of cues signals food (*10*, *11*). Second, animals need to discriminate sequences paired with food from those that are not. Rats can learn two-element sequences, A followed by B rather than B

Table 1. Illustration of the three rules and mean rate (per minute) of food tray entries in each group in response to both reinforced (RF) and nonreinforced (NRF) rules on the first and last block of training. Each group received food after one of the three rules (XYX, XXY, or YXX) and no food after the other two rules. Although no differences were found between the three groups or between reinforced and nonreinforced rules during the first two blocks of training trials [F values (2,12) < 1], by the last block of training trials the main effect of reinforced versus nonreinforced rule was statistically significant in all three groups [$F(1,12) = 8.21$, $P < 0.01$], although neither the main effect for the different groups or the interaction were statistically significant [F values (2,12) < 1] suggesting that the effect was similar with all three rules.

Block of learning trials	Group 1 Rule 1 (XYX) ABA, BAB		Group 2 Rule 2 (XXY) BBA, AAB		Group 3 Rule 3 (YXX) BAA, ABB	
	RF	NRF	RF	NRF	RF	NRF
First	12.66	12.58	9.14	9.81	13.94	13.97
Last	25.97	24.18	24.47	22.97	32.08	29.06

[1]Department of Psychology, University College London, Gower Street, London WC1E 6BT, UK. [2]Department of Education, University of Oxford, 15 Norham Gardens, Oxford OX2 6PY, UK.

*To whom correspondence should be addressed. E-mail: robin.murphy@ucl.ac.uk (R.A.M.); e.mondragon@ucl.ac.uk (E.M.)

followed by A (*12*), as signals for food. The third is the ability to transfer a rule to novel instances. We tested whether the rats learned a generalized rule rather than simply the trained instances. The transfer involved novel sequences that were either consistent or inconsistent with a previously trained rule but used stimuli that rats had never been exposed to or trained with.

Discrimination between stimuli is not a fixed ability, because animals usually show some generalization to novel items. For example, to the extent that a new stimulus C has more perceptual similarity to A than to B, animals will treat C like it is A (*10*, *13*). Furthermore, similarity is not restricted to perceptual features. Animals will treat two different cues with the same consequences (for instance, being paired with food) as similar (*14*). Finally, animals will sometimes treat stimuli that have the same relative position on a stimulus dimension (such as being greater than or brighter than) as similar (*15*). In our experiment, the stimulus changes within and between sequences were not monotonically distributed along a dimension. There were no stimulus-specific perceptual cues or common consequences that previous research might have predicted could have allowed generalization of behavior.

The rules were based on experiments that have shown that prelinguistic human infants behave as if they learn rules embedded in sequences created from auditory phonemic cues (*5*, *16*). We used patterns of visual (experiment 1) and auditory (experiment 2) stimuli that obeyed a similar rule.

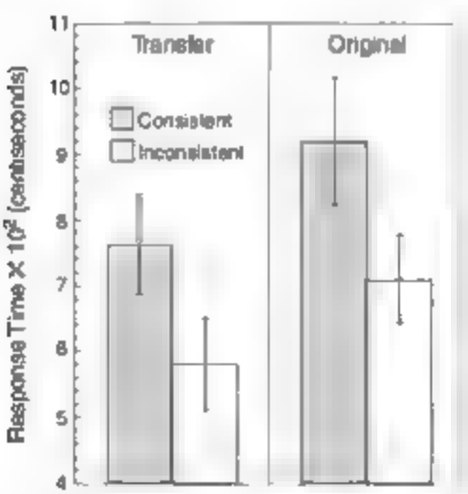


Fig. 1. Mean time (in centiseconds) with head in the food trough during the final 10 s of a sequence, subtracting the 10-s prestimulus interval for that trial. Response rates are shown separately for consistent sequences and inconsistent sequences, and the analysis confirmed that there was more responding to consistent than inconsistent sequences during the transfer test in the absence of food [$F(1,13) = 5.35$, $P < 0.05$] and during a test of the original stimuli [$F(1,13) = 4.87$, $P < 0.05$].

The cues in experiment 1 consisted of three-element sequences composed of short 10-s exposures to two visual cues: a bright light (A), or darkness (B) presented in a dimly lit environment. The rats were divided into three groups, with each group receiving Pavlovian conditioning for food after sequences that obeyed one of three rules. One group received food with the XYX rule (ABA and BAB), another group with the XXY rule (AAB and BBA), and finally a third group with the YXX rule (BAA and ABB). Each group also received trials with the other two sequences but without food. Would each group learn their rule for food and discriminate reinforced from nonreinforced sequences? We compared the rats' anticipatory responding for food during the third element of each sequence but before the food was available. The animals were able to discriminate among the sequences by the end of the experiment (Table 1). They came to respond more at the end of sequences that signaled food than to trials of either of the other two rules, and there was no evidence for differential responding to these two nonreinforced rules (*17*). Because A and B immediately preceded both food delivery and the absence of food, animals could not use the identity of the final stimulus to solve the discrimination. One possible solution involves the rats using unique pairs of stimuli that might have differentially signaled reinforced from nonreinforced sequences. Rats trained with XXY as the reinforced sequence might only have used the first pair of stimuli as a unique cue, whereas those trained with YXX might have used the final pair of stimuli. Finally, the group whose reinforced sequence was XYX could have learned to use a pair formed by the first and third stimuli. However, if animals did make use of different stimulus combinations in each group we would expect discrimination levels to differ because each of these discriminations would have involved quite different levels of difficulty. For instance, using a strategy based on the first two stimuli to discriminate XXY from the other patterns would imply that the third stimulus is uninformative. The third stimulus would effectively delay reinforcement, and weaken the discrimination, relative to reinforcement after YXX, in which food was contiguous with the unique stimulus pair. Overall, the notion that only stimulus pairs were crucial for learning does not account for the similar learning observed in all three groups. Discrimination performance did not differ among groups (*17*), suggesting that animals were anticipating food delivery by learning the whole triplet stimulus pattern. Rats had to use all three stimuli and their position in the sequence, a form of rule learning, to solve the discrimination.

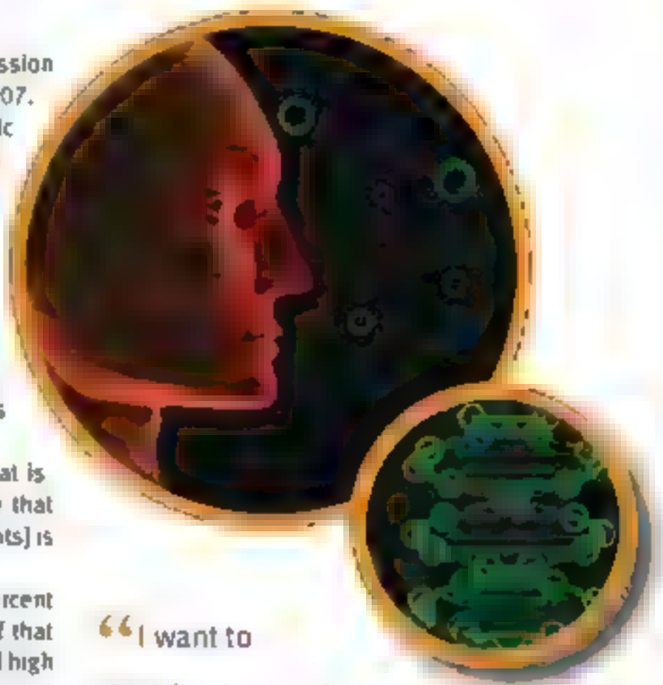
These results demonstrate that rats can learn three-element sequences, but it is possible that one method of solving the task involved memorizing the two sequences paired with food. Perhaps multiple instances of the rule (such as ABA and BAB) were treated as behaviorally similar but not perceived as examples from a broader category.

In a second experiment, we used a transfer test to explore this possibility. We tested whether rats could generalize learning of a rule to novel stimuli that were either consistent or inconsistent with the trained rule. We used a procedure similar to that in experiment 1, but because there were no differences in learning the three rules in experiment 1, we trained all subjects with rule 1 (XYX) using two auditory pure tones (A = 3.2 kHz and B = 9 kHz). Auditory cues were used because they allow a wider range of cues for the transfer test. Animals received food at the end of ABA and BAB sequences, but not after BBA, AAB, BAA, or ABB. After acquisition, we presented them with transfer stimuli composed of two novel pure tones (C = 12.5 kHz and D = 17.5 kHz). The stimuli were counterbalanced so that the stimuli in the roles of A, B and C, D were reversed for half of the animals and were chosen to ensure that no common frequency relation was present between the pairs. If rats had simply learned something specific about the reinforced elements ABA, they should have been unable to choose CDC and DCD over CCD, DDC, CDD, and DCC. The amount of time that the rats kept their heads in the food trough during the final element of the sequence was used as a measure of learning. The results of the transfer test are presented in Fig. 1, excluding two rats that failed to learn the initial discrimination. More anticipatory behavior for food was exhibited during sequences that were consistent with the previously learned rule, even though the rats had never been presented with these particular instances and there was no food presented during the test. As with experiment 1, animals responded more to the rule-consistent sequences than to both inconsistent sequences. There were no differences in responding to the two inconsistent rules (*17*). To test whether the rats remembered the trained sequences, we presented the original stimuli again, without any food reinforcement or any retraining, and found that they could still apply the rule to the original stimuli and consequently respond more to the previously reinforced sequences (Fig. 1).

The design of these experiments rules out a number of simple explanations of the rats' behavior. They could not have solved this discrimination simply by learning the final element of the sequences, because these elements were the same for both reinforced and nonreinforced sequences. Similarly, they could not use pairs of stimuli embedded in the sequences to solve the discrimination, because this would lead to differing levels of performance in experiment 1. The result of experiment 1 could be explained by the animals memorizing the sequential configuration of

GENOMIC BIOMARKER DISCOVERY:
BRINGING THE GENOME TO LIFE

A convergence of technological breakthroughs has taken genomic biomarker discovery to a new level. Between advances in sequencing, evolving array designs, and a more sophisticated understanding of genome architecture, simple tests for glucose, cholesterol, and human chorionic gonadotropin could soon be competing for pharmacy shelf space with gene expression– and epigenome-based diagnostics that promise to detect diseases earlier, stratify patients into treatment classes, and identify those most likely to respond to therapies. First to benefit cancer patients. The era of personalized medicine is just around the corner. **By Jeffrey M. Perkel**



The US Food and Drug Administration approved the first gene expression microarray–based test for use in the United States on February 6, 2007. MammaPrint, from Amsterdam-based **Agendia**, is a breast cancer diagnostic that uses the combined expression of 70 genes to predict whether a woman is at high or low risk for her cancer to recur.

Based on **Agilent Technologies**' 60-mer oligonucleotide microarrays and costing about $4,200, MammaPrint he ps doctors and their patients make more informed decisions, says René Bernards, Agendia's scientific director and professor of molecular carcinogenesis at the **Netherlands Cancer Institute (NCI)**.

Patients classified as having a high risk of metastasis have a 50 percent 10-year survival, says Bernards, versus 96 percent for low-risk patients. A combination of adjuvant chemotherapy and hormonal therapy, however, increases the high risk pool's survival to about 70 percent.

"If you can increase your chances of survival from 50 percent to 70 percent, that is fairly significant," he says. At the same time, Bernards continues, "you can show that the net benefit of adding chemotherapy to the hormonal therapy [for low-risk patients] is essentially zero."

Such prognostic clarity can pay both medical and economic dividends. About 90 percent of US breast cancer patients currently receive chemotherapy, Bernards estimates. If that number were reduced to 60 percent (the fraction of women whose tumors are labeled high risk), the hea thcare system could save millions. Meanwhile, patients who either do not need, or cannot benefit from, chemotherapy avoid its unpleasant side effects.

It truly is individualized medicine. And MammaPrint is just the beginning. Researchers in both academia and industry are scrambling to find new genomic biomarkers based on both gene expression and epigenetic modifications, especially for cancer.

Serving Different Needs

Genomic biomarkers come in several flavors, including disease detection and classification, treatment response prediction, treatment efficacy, and prognosis.

Their development follows a common theme. Typically, genomewide microarrays are used in a pre liminary screen on a relatively small set of cell lines or tissue biopsies. The resulting biomarker candidates are then validated on an independent sample set, and typically migrated to a different assay platform, such as quantitative real-time PCR, which can better probe a small number of loci for a large number of samples.

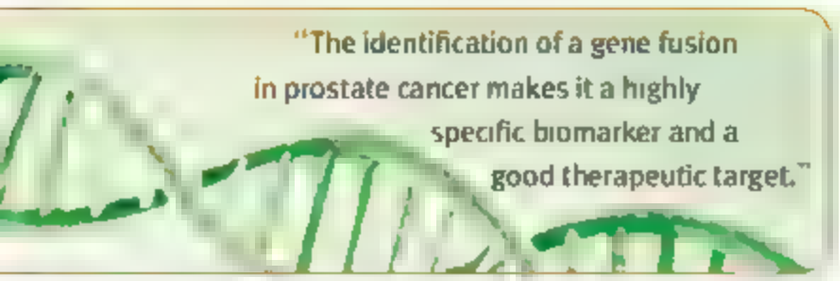
Certain groups at **Wyeth** identify candidate patient-selection biomarkers using **Affymetrix**'s U133 whole genome microarray and targeted signaling protein array analysis of drug-resistant and drug-sensitive cell lines, says Christina Coughlin, associate director of translational medicine efforts in the oncology pipeline at the company's Collegeville, Pennsylvania, facility. The 100 or 200 most-promising hits are then culled via more focused work using xenografts and genetically engineered mice.

"I want to identify who are the patients who will have true clinical benefit, and include them in my trial, and who will be resistant, and exclude them from the trial," says Coughlin.

Other biomarkers, like those found on MammaPrint's 70-gene panel, help

"I want to identify who are the patients who will have true clinical benefit, and include them in my trial, and who will be resistant, and exclude them from the trial."

Inclusion of companies in this article does not indicate endorsement by either AAAS or Science, nor is it meant to imply that their products or services are superior to those of other companies.



guide treatment decisions. Similarly, Oncotype DX, from Redwood City, California–based **Genomic Health**, uses the combined expression of just 21 genes to predict both a patient's breast cancer prognosis and her likelihood of responding to chemotherapy.

Steve Shak, Genomic Health's chief medical officer, explains that for every 100 women diagnosed with node-negative, ER-positive breast cancer, 85 "will survive without recurrence" with surgery and hormonal therapy alone. Of the 15 who will recur, chemotherapy benefits four. In other words, just 4 percent of breast cancer patients both need and can benefit from cytotoxic chemotherapy.

"So because we don't know who is more likely to recur, we 'punish' the many to benefit the few," he says.

Oncotype DX uses quantitative RT-PCR to produce a "recurrence score" between 0 and 100, with the likelihood of recurrence ranging from 4 percent for a low score to 35 percent for a high one. As with MammaPrint, its gene panel arose by retrospectively mining earlier studies correlating outcome with gene expression in breast cancer.

But key to Oncotype's development was a novel method to perform RT-PCR on RNA in formalin-fixed tissues, says Shak—an advance that took Genomic Health two years to pull off, but which opened up for the company all those archived tissue blocks left over from 30 years of clinical studies in the United States and abroad.

"The RNA was there, but it was fragmented. By optimizing the assay to look at those fragments of RNA, we were able to reliably quantify it," he says.

Expect the Unexpected

The success of Oncotype DX and MammaPrint highlights the continued utility of archived cancer array datasets. But how to locate, integrate, and analyze them all? Arul Chinnaiyan, professor of pathology and urology at the **University of Michigan School of Medicine**, and former graduate student Daniel Rhodes developed what has turned out to be an unexpectedly fruitful resource for researchers looking to do just that.

Oncomine (available through **Compendia Bioscience**), is an online database of gene expression in cancer. Its 608 million datapoints represent 342 studies on 40 different disease types —24,250 arrays in total.

According to Chinnaiyan, the project arose out of "mundane" necessity. "We were doing more gene expression analyses, and were always asked by academics whether their favorite genes were dysregulated in a specific cancer, and we had to mine the datasets to find out."

They developed Oncomine to "empower the average oncologist or biologist to ask the question themselves."

But Chinnaiyan quickly realized Oncomine was a useful hypothesis-generating tool, as well. "It's like a Google for high throughput gene expression data in cancer," he says.

His lab has plumbed Oncomine to discover several potential biomarkers, including a novel chromosome translocation (typically associated with "liquid cancers") in prostate cancer, a solid tumor. The translocation produces a gene fusion between TMPRSS2 and an Ets transcription factor, empowering Ets with androgen responsiveness, and represents a unique opportunity, he says.

"The identification of a gene fusion in prostate cancer makes it a highly specific biomarker and a good therapeutic target," he says, just as the Bcr-Abl fusion is both a biomarker for chronic myelogenous leukemia, and the target of its therapy, Gleevec.

So unexpected was the observation, Chinnaiyan recalls, he initially thought it must be wrong. Now San Diego–based **Gen-Probe** is developing a urine-based detection diagnostic for prostate cancer based on the discovery, he says.

Epigenetic Biomarkers

Not all genomic biomarkers are based on gene expression. Researchers have become increasingly aware recently of the critical role that epigenetic changes—whether involving DNA methylation or histone modifications—play in both development and disease.

Most epigenetic biomarker discovery focuses on DNA hypermethylation. One of the most common events in tumor formation, it turns out, is silencing tumor suppressors and cell cycle control genes by methylating their (otherwise unmethylated or partially methylated) promoters.

Kenneth Nephew, Professor of Cellular and Integrative Physiology at **Indiana University School of Medicine**, who studies epigenetic changes as indicators of breast and ovarian cancer, says such changes represent good disease biomarkers for three reasons: they are stable, common, and involve a gain of signal.

They also are easily detected. As tumor cells die they shed their DNA into the circulation. So, researchers are looking for ways to diagnose diseases earlier using the methylated DNA they find in blood, urine, feces, and sputum.

The process is essentially the same array-based method as with single nucleotide polymorphism and gene-expression biomarker discovery, says Art Petronis, of the **Centre for Addiction and Mental Health** in Toronto. "The difference is we have to use tiling arrays."

Rather than focusing on protein-coding or polymorphic DNA loci, tiling arrays cover the entire genome (or a large fraction of it). Petronis uses two in his search for epigenetic hallmarks of complex non-Mendelian disorders: a homemade array of some 12,000 CpG islands, and Affymetrix's higher density GeneChip Human Tiling 2.0R Array Set, which spans the entire human genome with 35-bp spacing on seven arrays.

The other difference, Petronis adds, "is we have to enrich the fraction that contains a high density or low density of methylcytosine, and then interrogate that fraction."

There are basically two ways to perform those steps, one based on differential restriction enzyme sensitivity of methylated DNA, the other on bisulfite conversion of methylcytosines to uracil.

Peter Laird, director of the **University of Southern California Epigenome Center**, uses two bisulfite conversion strategies in his search for colorectal cancer classification and ovarian cancer detection biomarkers. Laird starts with **Illumina's** Infinium assay, an array-based method that probes 27,000 CpG islands in a 12-sample format. He then validates those hits using a more sensitive technique called MethyLight, which is based on **Applied Biosystems'** TaqMan quantitative PCR chemistry and the ability of primers to distinguish minute sequence variations.

"With MethyLight we can detect 0.1 percent methylated versus 99.9 percent unmethylated DNA," he says. "With a profiling technology (like an array) you couldn't detect that."

CREDIT: ISTOCKPHOTO.COM/MEYANS

Featured Participants

Abbott Molecular
www.abbottmolecular.com

Affymetrix
www.affymetrix.com

Agendia
www.agendia.com

Agilent Technologies
www.agilent.com

Applied Biosystems
www.appliedbiosystems.com

Broad Institute of Harvard and MIT
www.broad.mit.edu

Centre for Addiction and Mental Health
www.camh.net

Cold Spring Harbor Laboratory
www.cshl.edu

Compendia Bioscience
www.compendiabio.com

Epigenomics
www.epigenomics.com

Genomic Health
www.genomichealth.com

Gen-Probe
www.gen-probe.com

Illumina
www.illumina.com

Indiana University School of Medicine
www.medicine.iu.edu

Netherlands Cancer Institute
www.onderzoekinformatie.nl/en/oi/

Nimblegen
www.nimblegen.com

Northwestern University
www.northwestern.edu

Oncomine
www.oncomine.org

Orion Genomics
www.oriongenomics.com

Sidney Kimmel Comprehensive Cancer Center
www.hopkinskimmelcancercenter.org

University of Michigan School of Medicine
www.med.umich.edu/medschool

University of Southern California
www.usc.edu

Wyeth
www.wyeth.com

According to Victor Levenson, research associate professor at **Northwestern University**, the discovery of bisulfite conversion chemistry was a "breakthrough" in epigenetic research. "Before that, it was extremely difficult to identify sequence-specific epigenetic modifications," he says.

"But the breakthrough came at a cost," he adds. First, the conversion usually destroys 85 percent to 95 percent of input DNA, creating potential bias. And PCR on the surviving DNA is difficult, both because the two strands are no longer complementary, and because the sequence has degenerated from a tetranucleotide alphabet to one with only three bases.

Instead, Levenson uses a restriction enzyme–based approach for biomarker discovery. So does biomarker firm **Epigenomics**, says Achim Plum, vice president of corporate communications. Epigenomics' process, called differential methylation hybridization (DMH), involves fragmenting genomic DNA, attaching primers to all fragment ends, and then digesting them with a methylation-sensitive enzyme. While methylated DNA will survive the treatment, he explains, unmethylated DNA will not. Thus, it is excluded from subsequent steps, in which the fragments are amplified and used to probe a custom Affymetrix array covering some 50,000 genomic fragments—mostly promoters and CpG islands.

"The beauty of the array is we put all our knowledge of what regions to look at on the array," says Plum. "So the fragments covered by probes are highly selected based on our knowledge and experience."

Plum says the company has used DMH to drive internal efforts to develop a lung cancer detection diagnostic. The process took less than a year, he says, from discovery to clinical proof-of-concept with a set of markers that could detect lung cancer with 69 percent sensitivity and 91 percent specificity in blood. Epigenomics' most advanced diagnostic, in development with **Abbott Molecular**, is a colorectal cancer–screening test using methylated Septin-9 DNA in blood as a biomarker.

Orion Genomics uses a similar strategy. Developed by **Cold Spring Harbor Laboratory**'s Robert Martienssen, Orion's MethylScope assay involves shearing genomic DNA to approximately 1-kb fragments (which enables complete genome coverage), splitting that pool into two fractions, digesting one with the methylation-dependent enzyme McrBC, size-fractionating the products to select for the uncut fragments, and hybridizing the two pools to a 2.1-million oligonucleotide array from **Nimblegen**.

With the right marker, even a single region can be informative, says Orion CEO Nathan Lakey. "We found even single-locus epigenetic changes that can discriminate breast cancer tumors from nontumor tissue with 90 percent sensitivity and 96 percent specificity."

Who Needs Arrays?

Bradley Bernstein, of the **Broad Institute of Harvard and MIT**, has developed a genomewide screening approach for histone modifications that eliminates DNA arrays.

"ChIP-Seq" combines chromatin immunoprecipitation (ChIP) and Illumina's next-generation sequencing technology, acquired from Solexa. Like ChIP-on-chip, which reads ChIP data using whole-genome microarrays, ChIP-Seq provides a genomewide view of chromatin changes. The difference, says Bernstein, is that ChIP-on-chip measures fluorescence intensity, while ChIP-Seq counts the absolute number of times a particular genomic fragment is detected.

"ChIP-on-chip is an analog readout, whereas sequencing is a digital readout," he explains.

This past August, Bernstein coauthored a study that used ChIP-Seq to map several histone modifications across the genome in mouse embryonic stem cells, neural progenitors, and embryonic fibroblasts. The data shed light on the cells' transcriptional past, present, and future, Bernstein said.

"It is a way to really precisely characterize the state of the cell," he explains. "The patterns tell you about what the cell is doing, and what it can become later."

A study, published February 17 in *Nature*, describes a similar method for probing methylated DNA with single-base resolution. Called BS-Seq, the technique combines Solexa sequencing technology with bisulfite conversion; it was used to scan some 93 percent of all possible methylation sites in the *Arabidopsis* genome.

Such data will be invaluable as the National Institutes of Health's $190 million Epigenome Project, part of the so-called "Roadmap Initiative," kicks off.

"The idea is to collect genomewide maps for some defined number of primary cell types and models, in order to understand the normal epigenome," Bernstein says. "Once you do that, you can begin to look at diseased epigenomes."

Yet Bernstein admits DNA methylation will likely be more valuable for biomarker development than histone modifications, because while tumors release DNA into the circulation, they do not shed intact chromatin. Thus, histone marks can only be assayed in intact cells and tissues.

"The information content in the histones and chromatin is much richer, but it's more difficult to assay," he says.

However these and other tests play out, says Stephen Baylin of the **Sidney Kimmel Comprehensive Cancer Center** in Baltimore, Maryland, the result will be a more individualized medical experience.

Take colon cancer, for instance. "Right now you do a blood test in the stool to screen for blood," he says. "That is suggestive, but not indicative of colon cancer. But if you can get the DNA, it will be much more diagnostic." After all, why have a colonoscopy if you don't need it?

Jeffrey Perkel is a freelance writer based in Pocatello, Idaho.

DOI: 10.1126/science.opms.p0800024

New Products

Gradient Thermal Cyclers

The new line of thermal and gradient thermal cyclers offers users the combination of outstanding flexibility and ease of use thanks to a host of standard features, yet the instruments are competitively priced and compact enough to suit most laboratories and budgets. They incorporate a large, easy-to-view control panel for intuitive, simple touch-screen programming of the entire protocol. The multiple program options and reaction formats ensure flexibility across a range of thermal cycling reactions and applications. They feature heated lids that effectively seal and reduce condensation in the sample vessel. The gradient model has four quick-change block options for 0.2-ml or 0.5-ml tubes, 96-well or 384-well plates, and slides. The TC32/80 is an economical "nongradient" model for use when microplates or gradient are not required. It includes a combi-block for 0.2-ml and 0.5-ml tubes.

Cleaver Scientific
For information 44 (0) 1788 565 300
www.cleaverscientific.com

Gel Documentation and Marker System

The GEBR DNA Gel Marker System incorporates an advanced gel documentation system with an automated marking system that optimizes imaging and image analysis workflow in life science laboratories and reduces exposure times for delicate DNA to less than 5 seconds. Once the image is captured and the bands of interest have been selected, the robotic arm marks all four corners of the selected band, sterilizing the tool between each band. The gel is imaged and marked in just a few minutes. The bands of interest are clearly visible with a white light box and can be excised without causing harm to the DNA or to the user. The system's state-of-the-art technology improves all aspects of DNA gel analysis and manipulation, providing safety, enhancing DNA integrity, and improving speed for increased throughput. The system provides a solution to the longstanding problem of how to excise DNA bands from agarose gels without damaging the DNA by exposure to ultraviolet light.

Norgren Systems
For information 401-996-8589
www.norgrensystems.com

Gel Imaging System

The BioSpectrum 800 Imaging System for accelerated, two-dimensional gel imaging and analysis is a fully motorized and automated light-tight darkroom with high-resolution, 8.3-megapixel image capture capability. The fast capture time of the BioSpectrum 800, coupled with the ease of use of Nonlinear Dynamics' SameSpots analysis software, accelerates and simplifies research lab workflow. The system accommodates a variety of gel sizes. It provides the ability to analyze enough replicates for true statistical power in only a matter of hours, a key to generating robust and reliable data from proteomics research.

UVP
For information 909-946-3197
www.uvp.com

DNA from Fixed Tissue Samples

The Repli-g FFPE Kit supports the genomic analysis of formalin-fixed paraffin-embedded (FFPE) tissue samples, eliminating the problems associated with DNA fragmentation and damage caused by formalin fixation. The kit makes use of a novel DNA processing reaction, allowing the advantages of multiple displacement amplification, a key technology in the amplification of genomic DNA, to be extended to highly degraded DNA samples derived from FFPE tissue. Isolation of sufficient genomic DNA from FFPE tissue is often difficult due to the low amount of DNA available. The Repli-g FFPE Kit allows amplification of precious sample material while maintaining locus representation, enabling unlimited downstream analyses to be performed without further purification. The kit provides uniform amplification resulting in scalable and standardized DNA yields. Following lysis of the tissue section, the DNA is processed using novel buffers and enzymes that ligate fragmented DNA. A two-hour amplification reaction typically yields 10 µg DNA, and an eight-hour incubation can yield 40 µg.

Qiagen
For information 240-686-7660
www.qiagen.com

PCR Inhibitor Removal

The PowerClean DNA Clean-up Kit can purify DNA containing polymerase chain reaction (PCR) inhibitors. Removing the inhibitors can help with DNA that will not amplify or gives a weak signal. The manufacturer also offers a line of Power DNA isolation kits that feature this unique, patent-pending PCR inhibitor removal technology.

Mo Bio Laboratories
For information 800-606-6246
www.mobio.com

Oncology Panel

TissueScan is a panel of normalized complementary DNA from cancer tissues at various stages of progression. The panel allows scientists to quickly obtain a gene expression profile or a single nucleotide polymorphism (SNP) profile across hundreds of cancer tissues using simple, real-time polymerase chain reaction with gene-specific or SNP-specific primers. This capability is especially useful for microarray data validation in which candidate biomarkers are identified by comparing a relatively small number of samples. The tissue samples used for each panel cover all disease stages and are accompanied by comprehensive pathology reports.

OriGene
For information 888-267-4436
www.origene.com

Electronically submit your new product description or product literature information! Go to www.sciencemag.org/products/newproducts.dtl for more information.

Newly offered instrumentation, apparatus, and laboratory materials of interest to researchers in all disciplines in academic, industrial, and governmental organizations are featured in this space. Emphasis is given to purpose, chief characteristics, and availability of products and materials. Endorsement by *Science* or AAAS of any products or materials mentioned is not implied. Additional information may be obtained from the manufacturer or supplier.

A BOOST FOR VACCINE RESEARCH

Tremendous strides have been made in eradicating infectious disease scourges such as smallpox and polio that once killed and crippled millions; still, about 15 million deaths—or about one third of all deaths annually—result from infectious diseases worldwide, according to the World Health Organization. Of those, nearly half involve children under the age of 5 years, predominantly in poorer countries. The ongoing hefty death toll, the pharmaceutical industry's increasing interest in the research and development of vaccines, and plentiful funding from multiple sources all combine to provide a range of opportunities for postdocs and graduate students in vaccine research. The field is high growth and, perhaps more important, the fruits of this work promise to have a real impact on the health of the world's population. **By Emma Hitt**

A recent estimate of the global vaccine business by Lehman Brothers predicts that the vaccine industry will grow at a rate of 18 percent a year to $30 billion by 2011—significantly faster than the average of 4.4 percent for the pharmaceutical industry overall. "Today we see more opportunity than ever to be gained from investing in vaccine research," says **Keith Gottesdiener**, vice president for vaccine and infectious disease clinical research with Merck. "The reasons for the growth of the industry include increased funding, from both governmental and private sources, as well as an increased scientific understanding of the immune system that will enable real breakthroughs in the field," he says. And increased growth translates into a stronger job market with more opportunities (see "Expansive Opportunities").

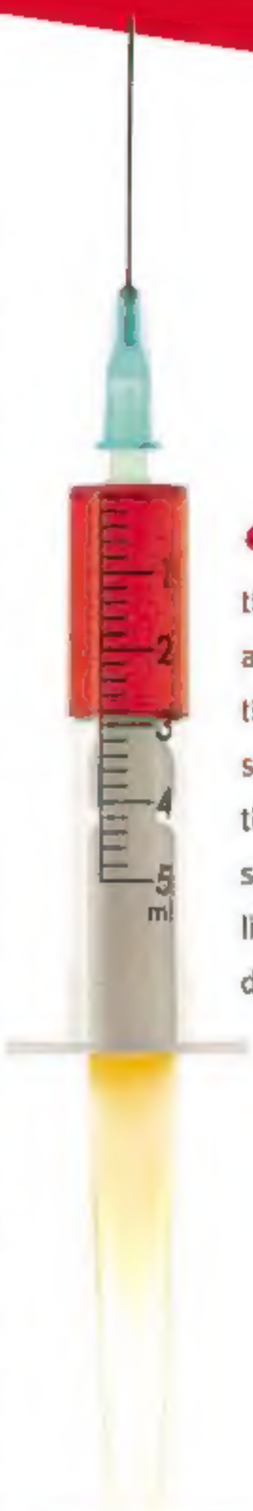
Recent blockbuster vaccines have included Gardasil and Cervarix, vaccines against certain human papillomavirus strains that cause cervical cancer, and Prevnar, the pneumococcal conjugate vaccine. Currently, about 20 vaccines are in development for HIV and are producing mixed results. The recent phase 2 STEP trial of the HIV vaccine (V520) showed disappointing results, but another vaccine made by GeoVax will be moving into phase 2 clinical trials in the next few months based on promising phase 1 results. VaxInnate is testing a universal flu vaccine that would work against all strains of the disease by using a Toll-like receptor (TLR) technology platform. Numerous experimental vaccines designed to ward off infections such as Japanese encephalitis virus, herpes simplex virus, and B cell lymphoma, and even to counter cocaine addiction, are under evaluation.

While vaccine research is clearly a growing field, this has not always been the case, notes **Alan Shaw**, CEO of VaxInnate, and the trend represents a stark contrast to the situation in the 1970s and 1980s, when many pharmaceutical companies exited the field due to poor economics and high risk of litigation. "A mere handful stayed in this business, but now, vaccines are seen as both economically and technologically interesting" he says. "A benefit of serving this area of research and participating in the development of a successful vaccine is that it is immensely satisfying. You can literally watch a disease disappear."

David Burt, who runs R&D in GlaxoSmithKline's North American hub, started working in the vaccine industry in 1989. "At the time, vaccines were thought of as being a low profit, low growth area but now pharmaceutical companies are viewing vaccines as a much more profitable area. New knowledge from the fields of immunology and microbiology has been applied to the development of vaccines in recent years. I think we have entered into a heyday for vaccines," he says.

Booster Shot for Funding

Increased funding has dramatically facilitated advances in the field within the past few years. To bring a vaccine through the R&D process to delivery can cost as much as $500 million, and the probability that a preclinical vaccine candidate will fail to come to market is about five times higher than its likelihood of success, according to a World Health Organization continued »



"A benefit of serving this area of research and participating in the development of a successful vaccine is that it is immensely satisfying. You can literally watch a disease disappear."

UPCOMING FEATURES

Translational Research (online only)—April 18
Diversity 2: Affinity Groups (online only)—May 9
Careers in Product Companies (online only)—May 23

CREDIT: PHOTO ILLUSTRATION: AMY HARDCASTLE (IMAGES: ISTOCKPHOTO.COM)

Vaccine Research and Production

"Our immunology program has a special interest in mucosal immunology and the development of improved clinical immunological tests that can be used in field evaluations in developing countries."
— John Clemens

(WHO) report. Since its inception in 2000, the Bill & Melinda Gates Foundation has contributed nearly $8.5 billion to global health, with HIV and tuberculosis (TB) receiving $2.29 billion and other infectious diseases about $1.96 billion. In the United States, the budget of the NIH's National Institute of Allergy and Infectious Diseases (NIAID) is about $4.5 billion for fiscal year 2008, and the WHO's Initiative for Vaccine Research budget was approximately $26.5 million for 2006-2007. "Funding and commercial support are increasing in the area of vaccine R&D, especially because of the high investment by entities such as the NIH and the Bill & Melinda Gates Foundation," noted **Marie-Paule Kieny**, director of the WHO's Initiative for Vaccine Research.

Global Collaborations

Vaccine research presents many opportunities for those interested in traveling to the far reaches of the globe. The GAVI Alliance (formerly known as the Global Alliance for Vaccines and Immunization) has been in existence since 2000 and seeks to "align public and private resources in a global effort to create greater access to the benefits of immunization." According to a 2008 UNICEF report, the Alliance's financial and programmatic support to vaccinate children around the world has helped prevent 2.9 million deaths since 2000 and decreased child mortality to less than 10 million per year for the first time ever.

"There has been a complete change in the attitudes of all involved in believing that vaccine coverage worldwide is possible and also that we can introduce and sustain that coverage in developing countries," notes **Rosamund Lewis**, with the GAVI Alliance. Lewis points out that they have been able to make relatively new vaccines, such as the pneumococcal vaccine, sold in the West since 2003, available to developing countries. "So the time lag between vaccine licensing and vaccine introduction in developing countries has been reduced from 20 years to less than 10 years for other products [Hib and Pneumo vaccines for example]." Another example is the rotavirus vaccine, which was put on the market in the West just last year. "We are already introducing this vaccine to developing countries, so we have reduced the gap to two years, which is our aim. The other aim is to introduce vaccines that are of special interest to developing countries, such as a malaria vaccine," she says.

The International Vaccine Institute, established under the Vienna Convention of 1969 with the signatures of 40 countries and the WHO, is headquartered in Seoul, South Korea. The IVI also focuses on bringing new vaccines to developing countries. According to **John Clemens**, the IVI's director-general, its immunology programs have focused on research and supply of vaccines against diarrheal diseases, bacterial meningitis and pneumonia, dengue fever, and Japanese encephalitis. "Our immunology program has a special interest in mucosal immunology and the development of improved clinical immunological tests that can be used in field evaluations in developing countries," he says. "A particularly noteworthy accomplishment by the IVI has been the development of a sublingually administered vaccine which has demonstrated efficacy in eliciting mucosal immunity against respiratory and genital tract infections," he notes.

Online Resources

Industry:

Crucell N.V.
www.crucell.com/Careers

GlaxoSmithKline
www.us.gsk.com/html/career/career-summer.html

MedImmune
www.medimmune.appone.com

Merck
www.merck.com/careers

Novartis
nibr.novartis.com/careers/internship

Sanofi-Pasteur
http://www.sanofipasteur.com

Wyeth
www.wyeth.com/careers
pharmaresearch/university/postdoctoral

Government:

All job opportunities
jobsearch.usajobs.opm.gov

Centers for Disease Control and Prevention (CDC)
www.cdc.gov/employment/
www.cdc.gov/cogh/training/IETA

National Institutes of Health/ National Institute of Allergy and Infectious Diseases
www3.niaid.nih.gov/labs/training

National Institutes of Health Postdoctoral Opportunities
www.training.nih.gov/student

www.training.nih.gov/webforms/postdoctoral/application/adIndex.aspx

Global:

European and Developing Countries Clinical Trials Partnership (EDCTP)
www.edctp.org

The Global Fund to fight AIDS, Tuberculosis and Malaria
www.theglobalfund.org/en/jobs

US Agency for International Development (Washington, D.C.)
www.usaid.gov/careers/applicant.html

World Health Organization (WHO)
www.who.int/employment/en/

According to Clemens, the IVI has a variety of programs and opportunities ranging from laboratory internships to pre- and postdoctoral student and fellowship opportunities. The institute also has shorter term training opportunities lasting a few months. The IVI has three divisions: the Division of Translational Research focuses on the treatment and prevention of pediatric dengue, Japanese encephalitis, respiratory pathogens, and rotavirus as well as "diseases of the most impoverished," including typhoid fever, cholera, and shigella; the Division of Laboratory Sciences focuses on developing novel formulations and delivery systems for vaccines; and the Technical Assistance and Technology Transfer Program provides assistance in manufacturing practices and regulation in developing countries.

The Aeras Global TB Vaccine Foundation, a nonprofit product development partnership working in collaboration with private, government, industrial, and academic institutions, focuses on developing new vaccines against TB and ensuring their availability to all who need them globally. In September 2007, Aeras received a $200 million grant from the Bill & Melinda Gates Foundation to accelerate TB vaccine development. "We seek to make TB vaccine regimens for infants and adolescents available and licensed in the next seven to nine years," according to **Yasir Skeiky**, chief scientific officer of Aeras. "The disease of TB is probably worse than it was decades ago in terms of the health impact of drug resistant strains and coinfection with HIV," he says, "and drug resistance is now extensive."

According to Skeiky, Aeras's state-of-the-art manufacturing plant in Rockville, Maryland, can produce ~200 million doses per year of the recombinant bacille Calmette-Guerin (BCG) vaccine against TB, enough to meet global need. However, the BCG vaccine was developed 86 years ago, and TB, with increasing drug resistance, now kills more than 1.5 million people each year, second only to HIV/AIDS as the world's most

CREDIT: INTERNATIONAL VACCINE INSTITUTE (IVI)

deadly infectious disease. Aeras's current pipeline includes six phase 1/2 vaccine candidates designed to combat multidrug-resistant (MDR) (resistant to first-line drugs) and extensively drug-resistant (XDR) TB (resistant to second-line drugs).

Lewis of GAVI points out that there is a need for people who can work toward enabling the delivery of vaccines to the people who need them. "Students interested in this area could enter into any number of fields in addition to immunology research, including public health, epidemiology, and health economics," she says.

Needs and Emerging Challenges

Pressing needs in vaccine research will ensure opportunities in the field for years to come. "Obviously, we still have HIV, tuberculosis, and malaria staring us in the face," says VaxInnate's Shaw. "Most people in the industry would say that the easy vaccines have been developed already," he says.

"Understanding the host response will be a critical area in vaccine research," says **George Kemble,** vice president of research and development for MedImmune Vaccines, recently purchased by AstraZeneca. According to Kemble, many of today's vaccines were created by relying on the natural immune response mechanisms. "Our ability to control these reactions and use them to form a response to the vaccine will be critical in designing the next generation of vaccines," Kemble says.

Bioterrorism is another important and well-funded area in vaccine research. Concerns for national and international security have supplied impetus for an influx of funding for biodefense vaccines and emerging infection vaccines in the past five years. The Centers for Disease Control and Prevention in Atlanta, Georgia, is a pioneer of research on organisms such as anthrax, botulism, plague, smallpox, and viral hemorrhagic fevers.

CREDIT: © MERCK AND COMPANY, INC., VACCINE PRODUCTION AT MERCK'S WEST POINT, PENNSYLVANIA FACILITY

"The reasons for the growth of the industry include increased funding, from both governmental and private sources, as well as an increased scientific understanding of the immune system that will enable real breakthroughs in the field."
— Keith Gottesdiener

Diverse Skill Sets

The need for skilled scientists at several levels is likely to increase in this field. "The field of vaccine research intersects with areas of public health, economics, politics, ethics, and patent/intellectual property issues. Opportunities exist for people with expertise including infectious diseases, epidemiology, molecular biology, immunology, and biostatistics," says **Eileen Barry,** with the Center for Vaccine Development at the University of Maryland in Baltimore.

Notably, the basic skills needed for vaccine research are taught in most universities: immunology, molecular biology, protein purification and analysis, epidemiology, biostatistics, biochemical engineering. "There will always be a need for people who have a strong background in basic science such as biology, chemistry, bacteriology, and immunology. Molecular biology and bioinformatics will be two of the areas in vaccine research with the strongest needs," says GSK's Burt.

So whatever your life science training, there could be a rewarding job for you in vaccine research. Exciting opportunities to travel and the ability to make a real difference to the health of humankind await.

Emma Hitt is a freelance medical and science writer residing in Marietta, Georgia.

DOI: 10.1126/science.opms.r0800051

Expansive Opportunities

A keyword search on "vaccine" at *Science* Careers (www.sciencecareers.org) and other science job boards illustrates a range of diverse opportunities in vaccine research for which a Ph.D. is preferred or required. Here are sample job titles and the key responsibilities:

Business Development Associate. Identify and secure partners for products and technology developed by the company. Responsible for securing strategic product development and commercialization partnerships in the context of obtaining structured and long-term revenue to support and sustain R&D/commercial programs.

Clinical Researcher. Ensure that clinical trials are conducted in a timely manner and in accordance with regulations. In collaboration with the global clinical team, responsible for overall assessment, initiation, and management of vaccine clinical research investigators and sites.

Director of Quality Assurance. Ensure that vaccine and diagnostics distribution activities are performed in compliance with the good distribution practice requirements and function as the point of contact for companies performing distribution activities for vaccine products.

Program Manager. Manage programmatic, scientific, and administrative aspects of grants research programs at a vaccine research institute.

Project Director. Manage a large consortium of investigators in imaging (PET/CT) of animal models of tuberculosis, for the purpose of understanding current chemotherapy as well as development of new therapies.

Pre- and Postdoctoral Opportunities in Biodefense and Emerging Infectious Diseases. Undergo research training in biodefense and infectious disease. Area of research might include bacterial, fungal, parasitic, and viral pathogenesis, infectious disease epidemiology, molecular diagnostics, therapeutics and drug design, and vaccine development.

Process Engineer. Assist with technical development of vaccines by providing direct process support for the manufacture of vaccine products. Requires B.S./M.S. in chemical/mechanical engineering or the biological sciences fields.

Regulatory Affairs Scientist. Guides and implements the regulatory process for preclinical evaluation of HIV vaccine products leading to phase 1 IND submissions to the US Food and Drug Administration within the US Military HIV Research Program.

Research Fellow, In Vivo Vaccine Research. Responsible for providing creative inputs on vaccine development strategies, design and construction of live recombinant viruses, and subsequent characterization. Additional postdoctoral experience may be required.

Research Specialist. Research specific vaccines for the development and evaluation of novel vaccine delivery systems.

Vaccine Formulation Development. Formulates development activities of prophylactic and/or therapeutic vaccine products from phase 1 clinical trials through commercialization. The position will involve use of innovative and modern techniques to develop and characterize viable dosage forms with appropriate stability profiles of candidate molecules to meet project needs.

Aeras Global TB Vaccine Foundation
www.aeras.org

AstraZeneca
www.astrazeneca.com

GAVI Alliance
www.gavialliance.org

GlaxoSmithKline
www.gsk.com

International Vaccine Institute
www.ivi.org

Merck
www.merck.com

University of Maryland Center for Vaccine Development
www.medschool.umaryland.edu

VaxInnate
www.vaxinnate.com

World Health Organization
www.who.int